http://www.scribd.com/document_downloads/18499113?extension=pdf
08/10/2009

Elías Mejía Mejía

TÉCNICAS E INSTRUMENTOS DE INVESTIGACIÓN

UNMSM

LIMA - 2005

ISBN: 9972-834-08-05

Hecho el Depósito Legal en la
Biblioteca Nacional del Perú N.º 2005-8142

Primera edición
Lima, noviembre de 2005

© Elías Mejía Mejía
© Unidad de Post Grado de la Facultad de Educación de la UNMSM
Serie: *Textos para la Maestría en Educación*

Diseño y diagramación:
Centro de Producción Editorial e Imprenta de la
Universidad Nacional Mayor de San Marcos

FACULTAD DE EDUCACIÓN
UNIDAD DE POST GRADO

Decana	:	Dra. Aurora Marrou Roldán
Director de la UPG	:	Dr. Elías Mejía Mejía
Comité Directivo de la UPG	:	Dr. Kenneth Delgado Santa Gadea
		Mg. Gonzalo Pacheco Lay

ÍNDICE

PRÓLOGO 09

CAPÍTULO I

INSTRUMENTOS DE ACOPIO DE DATOS

1. La medición 13
2. Postulados de la medición 14
3. Niveles de medición 15
4. Medición de variables conductuales 18
5. Instrumentos de acopio de datos 19
6. Instrumentos para el estudio de variables conductuales 20
7. Cualidades de los instrumentos de acopio de datos 22
8. Principios para construir pruebas 34
9. Técnica para elaborar pruebas 35
10. Indicadores para establecer la calidad de una prueba 39
11. Grado de dificultad de la prueba 40
12. Índice de discriminación de una prueba 41
13. Los ítemes 42
14. Clasificacion de los ítemes 42
15. Determinacion del grado de dificultad de un ítem 49
16. Índice de discriminación de un ítem 50

Lecturas complementarias

Fundamentos de medición

Donald Ary 55

Recolección de los datos

Roberto Hernández Sampieri 80

CAPÍTULO II

EL MUESTREO

1. Población y muestra 95
2. Determinación del tamaño de la muestra 99
3. Error de muestreo 102
4. Ventajas de la técnica de muestreo 103
5. Muestreo probalístico 104
6. Muestreo no probalístico 115

Lectura complementaria

Condiciones y elementos de las muestras
Restituto Sierra Bravo 119

CAPÍTULO III

EL PROCESO DE PRUEBA DE HIPÓTESIS

1. Hipótesis nulas y prueba inversa 139
2. Errores que se pueden cometer al adoptar decisiones 141
3. Prueba de hipótesis 142
4. Nivel de significación 142
5. Grados de libertad 143
6. Prueba chi cuadrada (x^2) 144
7. Análisis de varianza 148
7. Análisis de regresión múltiple 160
8. Prueba *t* de Student 167

Lectura complementaria

Prueba de hipótesis y error estándar
Fred Kerlinger 177

CAPÍTULO IV

EL INFORME CIENTÍFICO

1. Comunicación de la investigación 197
2. Extensión del informe 204
3. Estilo de la redacción del informe 204
4. Referencias bibliográficas 205

Lecturas complementarias

El informe de investigación
Orfelio León 211
Análisis, interpretación y comunicación de los resultados
Donald Ary 221

BIBLIOGRAFÍA 243

PRÓLOGO

Las técnicas para la realización de investigaciones científicas son muy variadas y tienen distintos propósitos, pero todas ellas resultarán siendo inútiles si antes no se ha comprendido a cabalidad la metodología de la ciencia para producir conocimientos. Esta ardua tarea requiere formación sólida y lecturas intensas para comprender la racionalidad de la ciencia, explicarnos los hechos a partir de hipótesis y tomar decisiones con respecto a ellas en función de la evidencia que se halle en la observación de los hechos. Nada tan evidente: si los hechos se presentan en el sentido propuesto por las hipótesis, éstas se aceptan; mientras que si los hechos no se presentan en la forma planteada por los hechos, éstas se rechazan. En otras palabras, son los hechos, y no las conjeturas, los que dirimen acerca de la validez del conocimiento, pues sería de una gran ingenuidad pretender dar por válidas las hipótesis que no sean sustentadas por los hechos. Si una hipótesis no concuerda con los hechos, debe ser rechazada. No es posible lo contrario: rechazar los hechos para mantener la vigencia de nuestra hipótesis.

Este volumen complementa lo dicho en *Metodología de la Investigación Científica*. Aquí se proponen las técnicas más necesarias para realizar el trabajo de campo que demanda toda investigación.

Se explica lo relacionado con el proceso de construcción de los instrumentos para el acopio de datos. Para ello existen determinadas técnicas, pero si optamos por los métodos cuantitativos, debemos tener en cuenta que los fenómenos que vamos a estudiar, si cuando nos referimos a ellos, podemos hablar en términos de más o de menos, es decir, frente a dos fenómenos podemos decir que uno es más que otro, entonces es posible medirlos. Reconocer la magnitud de los fenómenos es requisito previo para poder medirlos. El investigador intuye que el fenómeno que estudia posee una determinada magnitud, pero le resulta difícil conocerla, medirla, porque no dispone de los instrumentos que le permitan acercarlo a la real magnitud de tales fenómenos.

También se dan las pautas para hacer el muestreo y se enfatiza que toda muestra, si partimos del razonamiento de que toda muestra es el subconjunto de la población, es representativa, y si no es representativa la supuesta muestra, sencillamente no es muestra.

En otro capítulo se explica el proceso de prueba de hipótesis con procedimientos estadísticos. Para ello se han adoptado como ejemplos algunas investigaciones realizadas en la Facultad de Educación, y que han servido a sus autores para optar los grados académicos correspondientes.

Finalmente, en el presente volumen se explica lo relacionado con el Informe de Investigación, que en realidad es la tesis que los graduandos deben redactar. La tesis es eso, el informe del trabajo de investigación realizado. El graduando debe realizar un proceso de investigación, debe recoger evidencias que contradigan o confirmen sus hipótesis y debe analizar sus resultados y tomar decisiones con respecto a ellas. Al final de todo este esfuerzo, el investigador, ya culminado el proceso de investigación, se recluye en su gabinete para dar cuenta, por escrito, del proceso de investigación que ha seguido y de la forma cómo ha sometido a contraste sus hipótesis con los hechos. Es decir, redacta su informe científico, empleando las formas del lenguaje más apropiadas al caso. Esta es la tesis que los estudiantes del programa de Maestría en Educación deben redactar, como resultado de un proceso de investigación realizado.

Se pretende, con este volumen, cubrir los vacíos que puedan existir en la aventura de hacer ciencia y reiteramos nuestra invitación a los participantes del programa para que asuman el reto de hacer investigaciones, informar acerca de sus hallazgos y, finalmente, optar el grado académico correspondiente, auténtico motivo de su incorporación a San Marcos de todos quienes estudian maestría.

Dr. Elías Mejía Mejía

Director de la Unidad de Post Grado

CAPÍTULO I

INSTRUMENTOS DE ACOPIO DE DATOS

INSTRUMENTOS DE ACOPIO DE DATOS

Hasta aquí se ha visto cómo el investigador busca la mejor estrategia para probar sus hipótesis. Pero es el caso que el investigador también necesita ciertas herramientas o instrumentos que le permitan recopilar datos de la realidad para probar sus hipótesis, pues no sería conveniente que los recopilara en base a sus cálculos, intuiciones o, como se dice, 'a ojo de buen cubero'. Los instrumentos que requiere el investigador pueden ser de diferentes tipos: de medición, de constatación, de acopio de información, de verificación de situaciones, etc. Los instrumentos más conocidos y los que proporcionan información más valiosa al investigador son los instrumentos de medición. La medición es una actividad muy importante cuando se trata de conocer la naturaleza de los fenómenos que proporcionan información precisa acerca de sus características.

1. LA MEDICIÓN

La medición es un proceso que consiste en asignar numerales a determinados fenómenos o eventos, siguiendo reglas previamente establecidas. Esta definición planteada por S. S. Stevens es intencionalmente muy genérica, pero es muy útil porque hace posible abarcar todos los aspectos que se derivan del proceso de medición y, además, porque permite sostener que es posible, teóricamente, medir cualquier fenómeno siempre y cuando las reglas tengan un fundamento racional o lógico.

Los fenómenos o eventos a los que se hace referencia son las variables, es decir, fenómenos que varían cuando asumen dos o más valores. En la investigación interesa examinar y analizar cada uno de estos valores.

Para realizar un proceso de medición es necesario reconocer que el fenómeno a medir tiene su propia magnitud y que el problema radica en que el investigador, con los instrumentos que dispone, no la puede conocer plenamente. La medición es, en estricto sentido, conocer la verdadera magnitud del fenómeno, de ahí que resulta muy importante que toda auténtica medición sea isomórfica con la realidad que se está midiendo, es decir, que los datos que se obtengan como resultado de la medición deben ser parecidos, equivalentes, o correspondientes a los que realmente posee el

fenómeno que se mide; aunque en realidad no se mide el fenómeno directamente, sino los indicadores de sus características. Es muy importante tener en cuenta esto cuando se realizan procesos de medición de fenómenos o variables del comportamiento que, por naturaleza, son muy elusivas, impredecibles y difíciles de identificar, como la *creatividad*, el *talento*, el *coeficiente intelectual*, la *personalidad*, la *agresividad*, el *rendimiento académico*, la *angustia*, etc. Como estos fenómenos ofrecen dificultades a los esfuerzos por medirlos, solo se puede inferir sus características a partir de la observación o el análisis de sus indicadores.

Analizar los aspectos relacionados con la medición es muy importante cuando se trata especialmente de la investigación cuantitativa en la que, en todo momento, se trata de emitir juicios de valor acerca de los fenómenos que se estudian y se trata de estimar, ponderar, puntuar, graduar, precisar, etc., sus características.

Para realizar procesos de medición es necesario que las variables que se estudian varíen en términos de más o de menos cantidad, es decir, que de una variable se pueda decir que la característica se presenta en términos de más cantidad o de menos cantidad. Por ejemplo, tratándose de la *inteligencia*, es posible decir que algunas personas tienen más inteligencia y otras tienen menos inteligencia. Si esto se puede decir de la *inteligencia*, entonces es posible medirla.

2. POSTULADOS DE LA MEDICIÓN

Los postulados de la medición son los referentes teóricos que fundamentan los procesos de medición. El investigador debe tenerlos muy presente si pretende realizar correctamente los procesos de medición. Estos postulados son los siguientes:

A es igual a B o A es diferente a B, pero no ambas situaciones a la vez. Este postulado es importante para la clasificación, porque cuando se clasifica se ubican, exhaustiva y excluyentemente, todos los elementos del conjunto que se está clasificando en una u otra categoría. Ningún elemento puede estar a medias en una categoría ni puede estar en más de una categoría a la vez. Un objeto puede ser igual ó diferente de otro, pero no puede ser igual y diferente a otro al mismo tiempo.

Si A es igual a B y B es igual a C, entonces, A es igual a C. Este postulado permite establecer la igualdad de los miembros de un conjunto en base a una característica común a partir de la comparación de los objetos.

Si A es mayor que B y B es mayor que C, entonces, A es mayor que C. Las relaciones también pueden ser "menor que", o "se halla a mayor distancia que", "es más fuerte que", "precede", "domina", etc. En este postulado se fundan la mayoría de las mediciones psicológicas y pedagógicas, pues al elaborar escalas, las categorías de éstas surgen por comparación con las características que poseen otros sujetos.

3. NIVELES DE MEDICIÓN

Los conceptos anteriormente expuestos permiten establecer una idea más aproximada de lo que en realidad es la medición. Sin embargo, todavía es necesario explicar lo relacionado con los niveles de la medición, concepto que también se debe a Stevens. Los niveles de la medición son las diferentes estrategias que permiten medir, con más o menos exactitud, el fenómeno. Estos niveles son cuatro: nominal, ordinal, de intervalo y de razón o proporcional.

3.1 Nivel nominal

Como su nombre lo indica, este nivel de medición consiste en asignar nombres o denominaciones a los sujetos o fenómenos de la realidad. Por ejemplo, todas las personas tienen dos nombres y dos apellidos. En nuestra sociedad el apellido paterno se lleva antepuesto al materno. Ésta es la regla pre establecida y, en este caso, se cumple la definición de medición. Ejemplos de situaciones que se miden en el nivel nominal abundan en la investigación de la conducta. Así, los estudiantes universitarios tienen su respectivo código de matrícula que los identifica; todos los ciudadanos están identificados por el número de su Documento Nacional de Identidad; los estudiantes de una sala de clases están identificados por el número de orden. Se puede codificar las categorías con 1 ó 2 para referirse al sexo masculino o al femenino, respectivamente. La misma codificación se puede usar para registrar si un estudiante está aprobado o desaprobado, si su padre vive o no, si es nacional o extranjero, si es rico o pobre, si proviene de Lima o de provincias, etc. En este caso, las variables se miden nominalmente al asignarles un código a cada uno de sus valores. Otros ejemplos en que la medición se realiza en el nivel nominal es cuando los jugadores de un equipo deportivo llevan en su camiseta un número que los identifica; los autos de carrera se identifican por sus números, así como los caballos de carrera o las candidatas en un concurso de belleza. En todos estos casos, los números no tienen el significado de cantidad. A quien se asigne el número 2 no significa que sea más con respecto a quien posea el número 1, sino que estas denominaciones sólo son 'nombres', 'rótulos', 'códigos' convencionales. La asignación de numerales se acepta como un proceso de medición nominal si previamente se han observado las reglas pre establecidas.

Las variables que se miden en este nivel generalmente son las categóricas. La investigación cualitativa se basa en este tipo de medición, al ubicar la característica que se estudia en una, y sólo en una, categoría. Estas categorías pueden ser 'nominadas' ó 'denominadas' según el libre albedrío del investigador; así por ejemplo, los valores del *estado civil* se pueden denominar como: *soltero*, *casado*, *viudo*, *divorciado*, *conviviente*, etc. El *tipo de gestión de las universidades* puede ser: *nacional* o *particular*. El *lugar de nacimiento* puede ser: *nacional* o *extranjero*.

Como se ha dicho, estas categorías son excluyentes y no puede darse el caso de alguien que sea *soltero* o *casado* a la vez, como tampoco puede ocurrir que alguien sea *nacional* o *extranjero*, al mismo tiempo.

3.2 Nivel ordinal

Como se habrá notado, la medición nominal sólo identifica a los sujetos. No proporciona ninguna información adicional. En cambio, la medición ordinal incorpora un elemento nuevo: indica el orden, precedencia o prelación en el que se hallan los sujetos. En este caso funciona el tercer postulado (*A* es mayor que *B* y *B* es mayor que *C*, en consecuencia *A* es mayor que *C*). Al emplear el nivel ordinal de medición se puede afirmar, por ejemplo, que en una carrera de caballos, el Nº 5 llegó 1º, el Nº 3, llegó 2º y que el Nº 7, llegó 3º. El orden de mérito de los estudiantes es un caso típico de medición ordinal, así como lo son los resultados de los exámenes de admisión a las universidades. Este nivel informa que, por ejemplo, ocupar el 5° puesto entre 300 alumnos es más meritorio que ocupar el mismo 5° puesto entre 6 alumnos, pero no informa con respecto a cuál de los estudiantes es mejor. El siguiente gráfico ilustra lo dicho:

Gráfico Nº 1

Un caso típico de medición en el nivel ordinal es la escala de evaluación que se emplea en los procesos de sustentación de tesis, en los que se evalúa como *aprobado* o *desaprobado*. Y dentro de la categoría de aprobado se distingue todavía las siguientes sub categorías: *aprobado por mayoría*, *aprobado por unanimidad*, *sobresaliente*, o *sobresaliente con recomendación de publicación de la tesis*. También los resultados de un proceso de admisión para ingresar en una Universidad se expresan en el nivel de medición ordinal, pues se dice que alguien ocupó el primer puesto, el segundo puesto, o el tercer puesto, etc. La medición ordinal requiere de la información que proporciona la medición nominal.

3.3 Nivel de intervalo

Como quiera que la medición ordinal proporciona información acerca de la precedencia, prelación o del orden en que se ubican los sujetos, se observa en muchos casos que el segundo lugar está muy próximo al primero y que el tercero,

muy lejos de los dos primeros, tal como pude verse en el gráfico N° 1. Para evitar estas distorsiones y hacer una medición más exacta, se emplea la medición en el nivel de intervalo, en el que se establecen distancias iguales para cada puntuación, es decir, se diseña una escala con intervalos iguales.

Las escalas intervalares o de intervalos iguales poseen las características de las escalas nominales y de las ordinales. La diferencia está en que las distancias de cada intervalo son iguales. La representación de este tipo de escala es la siguiente:

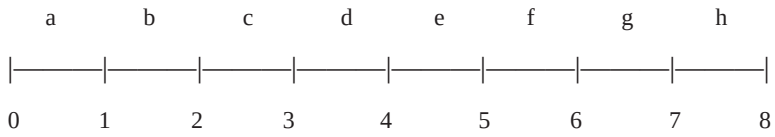


Gráfico N° 2

Por ejemplo, si se midieran cuatro objetos en el nivel intervalar se obtendrían los siguientes valores: 8, 6, 5 y 3. En este caso se puede afirmar, con toda razón, que la diferencia entre el primer objeto y el tercero (8 – 5 = 3) es igual a la diferencia que existe entre el segundo y el cuarto (6 – 3 = 3).

En la medición intervalar, los intervalos se pueden sumar o restar. Usando la representación anterior, se puede afirmar que el intervalo entre *c* y *a* es 3 – 1, ó sea 2; el intervalo que existe entre *d* y *c* es 4 – 3, ó sea 1. Además se pueden sumar los intervalos: 2 + 1 = 3

Y se puede comparar la distancia entre *d* y *a* (4 – 1 = 3) y la distancia entre *g* y *d* (7 – 4 = 3) y afirmar que las distancias son iguales. Lo que no se puede es afirmar que el aprovechamiento de *d* es dos veces superior al de *b*. Para formular este tipo de afirmación se requiere un nivel más elevado de medición. Con la medición intervalar se puede elaborar escalas, como la escala vigesimal para medir el *rendimiento académico*. Esta escala abarca del 1 al 20. La medición del coeficiente intelectual también se realiza en una escala de intervalo. La medición en el nivel de intervalo supone la medición en los niveles previos, es decir, en el nivel nominal y ordinal.

3.4 Nivel proporcional o de razón

Este nivel es el más elevado de la medición. Es el ideal de la medición científica. Una escala de este tipo, además de poseer las características de los niveles que se han descrito, parte del concepto de cero. Sin embargo, en algunos casos, la ubicación

del cero es relativa, lo que origina diferencias entre las escalas: la escala Celsius es diferente a la escala Farenheit, porque para medir la temperatura ambas escalas ubican el cero en posiciones diferentes. En el nivel de medición de razón se puede elaborar escalas que consideren valores sobre cero o bajo cero, como es el caso de las escalas para medir la temperatura o la presión atmosférica. Con estas escalas se puede realizar todas las operaciones aritméticas, como son la multiplicación y la división, además de la suma y la resta. Si existiera una escala para medir el *rendimiento académico* en el nivel proporcional, se podría decir que un estudiante, cuya calificación fuese 16, tendría un *rendimiento académico* dos veces superior con respecto de otro estudiante que obtenga la nota 08, sin embargo esto no es así en la realidad.

La medición de la temperatura o de la presión atmosférica son ejemplos de mediciones realizadas en el nivel proporcional, debido a que informan de temperaturas o niveles de presión atmosférica sobre cero o bajo cero, pero la medición que se expresa bajo cero, indica la presencia de alguna magnitud y no la ausencia total de la característica.

En las ciencias naturales, últimamente se está trabajando en el concepto del cero absoluto o de la ausencia total de la característica. Lord Kelvin considera que el punto donde no hay choques de moléculas que crean calor es el punto en el que se ubica el cero absoluto, ausencia de temperatura, y este punto es equivalente a – 273 grados Celsius, punto en el que los científicos consideran que no existe temperatura debido a que no existe actividad molecular. La medición de la temperatura, empleando este criterio, sería una medición exacta, pues el punto en el que se ubica el cero es precisamente el punto donde no existe temperatura. Sin embargo, parece que son muy pocos los sistemas de medición que parten del cero absoluto que permiten una medición exacta de los fenómenos.

4. MEDICIÓN DE VARIABLES CONDUCTUALES

Cabe enfatizar que los niveles de medición que se acaban de describir son jerárquicamente secuenciales, es decir, en el nivel siguiente se reflejan las características del nivel previo. Así, en la medición ordinal también están presentes las características del nivel nominal; en el nivel de intervalo se presentan también las características de los niveles nominal y ordinal y, en el nivel proporcional o de razón se presentan, como se ha dicho, las características de todos los niveles anteriores. Por ejemplo, sean 16,54 y 12,67 las puntuaciones halladas en la variable éxito académico de un grupo experimental y de otro de control, respectivamente. Tales puntuaciones han sido establecidas en el nivel de intervalo, pero pudieran también haber sido obtenidas en el nivel ordinal o nominal. Los datos a los que se

hace referencia informan que el éxito académico del grupo experimental es mayor que el éxito académico del grupo de control: nivel ordinal. Los mismos datos nos informan que el grupo experimental y el grupo de control son diferentes, pues están identificados con distintas puntuaciones: nivel nominal.

Para medir con mayor exactitud una variable es recomendable identificarla con la mayor precisión posible y, si fuera el caso, operacionalizarla, es decir, expresarla en función de sus indicadores o manifestaciones más significativas. También es importante destacar que la mayoría de las variables conductuales o las variables que estudian las ciencias sociales, en general, se miden en el nivel nominal, aunque algunas se miden en el nivel ordinal y muy pocas en el nivel de intervalo.

Cuando se mide variables en el nivel de intervalo, se elaboran escalas *ad hoc*, que tienen su punto de partida en la definición operacional de la variable que previamente debe haber elaborado el investigador. Pero se debe tener cuidado al analizar la información que proporciona este tipo de medición pues, en la mayoría de los casos, el nivel de intervalo sólo proporciona algo más de información de la que proporciona el nivel ordinal. Por ejemplo, si un estudiante obtiene un coeficiente intelectual de 110, esto no quiere decir que este estudiante sea 7 puntos más inteligente que un estudiante que, sea el caso, obtiene un coeficiente intelectual de 103.

Últimamente, los niveles de medición descritos por Stevens están recibiendo duras críticas de parte de los científicos sociales quienes consideran que estos niveles son muy rígidos y no reflejan, de manera natural, los fenómenos que suceden en el mundo real. Estos críticos consideran que algunas variables conductuales no necesariamente deben ser medidas en uno de estos cuatro niveles, como por ejemplo la *inteligencia* y el *aprendizaje*. Por ejemplo, estas variables no podrían ser medidas en el nivel de razón, porque carecen del cero, debido a que una persona no puede tener cero, es decir, ausencia total de inteligencia o de conocimientos, como se pretende cuando se dice que el aprendizaje se mide en la escala vigesimal de 0 a 20 puntos. No es posible asignar a un alumno el calificativo de cero porque un ser humano no puede tener la mente en blanco o vacía, que es lo que corresponde al concepto de cero. Este estudiante, por muy pocos conocimientos que posea, tendría alguna cantidad de conocimientos, sus conocimientos deben tener alguna magnitud; es imposible que no tenga conocimientos o esté con la mente vacía. El rendimiento académico no puede medirse a partir de cero.

5. INSTRUMENTOS DE ACOPIO DE DATOS

Con el nombre genérico de instrumentos de acopio de datos se denomina a todos los instrumentos que pueden servir para medir las variables, recopilar información con respecto a ellas o simplemente observar su comportamiento.

Los instrumentos que pueden medir las características de las variables se denominan tests o pruebas, son los instrumentos que sirven para medir distintas variables conductuales, en especial los resultados del aprendizaje. A través de los datos que proporcionan los instrumentos se trata de obtener información exacta sobre el logro de los aprendizajes y se detectan los éxitos y fracasos.

Cuando el investigador no puede medir directamente las variables debe recopilar información acerca de los fenómenos que le interesa conocer haciendo uso de ciertos instrumentos. Para ello emplea listas de cotejo, hace el análisis documental, construye escalas de opinión, etc. En otros casos, el investigador debe observar el comportamiento de las variables y entonces empleará instrumentos o guías de observación que, a su vez, pueden ser estructurados o no estructurados.

6. INSTRUMENTOS PARA EL ESTUDIO DE VARIABLES CONDUCTUALES

Las pruebas o instrumentos para estudiar variables conductuales son los siguientes:

- De medición de los aprendizajes.
- De actitudes o de opiniones.
- De inventario.

6.1 Pruebas de medición de los aprendizajes

Las pruebas para medir aprendizajes se deben elaborar pensando en las características y las necesidades de la investigación, por lo que su aplicabilidad se halla limitada a éstas y condicionada para la muestra elegida. La calificación de estas pruebas puede hacerse por la vía de dos métodos:

a) En referencia a una **norma**, y

b) En referencia a un **criterio**.

Se dice que una prueba está **referida a una norma**, cuando para establecer la significatividad de sus resultados, se recurre a la comparación del puntaje de un estudiante con el promedio alcanzado por el grupo al que pertenece este estudiante. Pruebas referidas a una norma son las que se aplican en los concursos de admisión a las universidades. En este caso, si un postulante alcanza 150 puntos en una prueba que tiene el puntaje máximo de 300 puntos y el promedio del grupo es 137, se puede decir que es un buen estudiante, pues se halla por encima de la norma del grupo. También se puede decir que este alumno ocupó el segundo puesto si sólo existiera un solo postulante que tenga un puntaje mayor que él y obviamente, que ingresó en la universidad.

Se dice que una prueba está **referida a un criterio** si previamente se ha establecido este criterio o nivel de desempeño específico para esta prueba. Por ejemplo, cuando se mide aprendizajes se usa, por lo general, la escala vigesimal y se establece que para estar aprobado es necesario obtener la nota 11. Este criterio cambia en otras circunstancias, por ejemplo, en las escuelas de postgrado de las universidades, cuando se establece que la nota aprobatoria es 13 ó 14, dentro de la misma escala vigesimal. Como se podrá apreciar, aquí aparece otro criterio, naturalmente más exigente. Como consecuencia de la aplicación de este criterio se establece una partición del grupo de estudiantes que han dado la prueba: aprobados y desaprobados.

6.2 Pruebas de actitud

En las pruebas de actitud, llamadas también de opiniones, se indaga acerca de las actitudes u opiniones de los individuos con respecto a ciertas situaciones que plantea el investigador. Por ejemplo, si se desea estudiar las actitudes u opiniones que asumen los docentes universitarios con respecto al tipo de formación profesional que se debe realizar en las universidades, el investigador planteará algunos reactivos acerca de los cuales deben opinar quienes resuelven esta prueba.

El formato básico según el cual se elaboran estas pruebas es el siguiente:

a) Se establece, con la mayor precisión posible, el nombre de la prueba.

b) Se redacta un enunciado con las instrucciones según las cuales debe actuar la persona a quien se aplica la prueba. Según las necesidades del investigador, la prueba puede ser nominal o anónima.

c) Se traza una tabla con tantas columnas como datos se pretende recoger y con tantas líneas como ítemes o reactivos se haya previsto plantear. En la primera columna se colocan los ítemes o reactivos que en realidad son proposiciones redactadas en sentido positivo o negativo a las que el examinado pueda responder emitiendo su opinión a favor o en contra. En las columnas siguientes se colocan las posibilidades de respuesta, organizadas en una especie de escala de tres, cuatro, cinco o seis alternativas. Suponiendo que la escala fuese de cinco valores, las alternativas pueden ser las siguientes: MA (muy de acuerdo); A (de acuerdo); I (indiferente); D (en desacuerdo); MD (muy en desacuerdo). El investigador debe asignar, en seguida, puntajes a cada tipo de respuesta, por ejemplo: MA = 5; A = 4; I = 0 (por ser indiferente); D = 2; MD = 1. En el Cuadro N° 1 se grafica lo que se acaba de decir:

FRAGMENTO DE UNA PRUEBA DE ACTITUDES

N°	OPINIONES	MA	A	I	D	MD
1	Es muy acertada la orientación que se ha adoptado en la Universidad para formar profesionales en el aspecto intelectual, fundamentalmente.	1	2	0	4	5
2	Todo buen profesional debe poseer habilidades para comunicarse con eficiencia con sus semejantes. Por eso son muy importantes los cursos de Lenguaje y Comunicación.	5	4	0	2	1
3	La identificación con la cultura y la realidad de su país es una condición básica para un exitoso desempeño profesional.	5	4	0	2	1
4	En la Universidad no deben llevarse a cabo actividades para orientar la conducta ética y moral de los alumnos, porque esos aspectos no son parte de la formación profesional.	1	2	0	4	5
5	Por lo general, los profesionales que, además de su especialidad profesional, conocen el arte y saben apreciarlo, no tienen tanto éxito en su especialidad profesional.	1	2	0	4	5

Cuadro N° 1

Para evaluar la prueba, el investigador debe establecer, previamente cuáles respuestas debe evaluar en sentido positivo y cuáles en sentido negativo. Las puntuaciones obtenidas las debe totalizar mediante una suma algebraica, es decir, sumando las respuestas en positivo y restando las respuestas en negativo.

7. CUALIDADES DE LOS INSTRUMENTOS DE ACOPIO DE DATOS

El éxito de una investigación, además de la plena identificación de las variables, la correcta formulación de la hipótesis, la estrategia adecuada para probar dichas hipótesis, etc., depende de la calidad de los instrumentos de acopio de datos, de ahí que resulta muy importante que el investigador asuma esta tarea con la mayor responsabilidad posible. Esto quiere decir que el investigador debe preocuparse en cuidar que los instrumentos que elabora para acopiar los datos, posean cualidades básicas y necesarias. Si un instrumento no posee las cualidades necesarias, la investigación puede adoptar un sesgo peligroso y se puede llegar a lo que los expertos denominan, resultados espúreos o incorrectos en una investigación.

Los instrumentos de acopio de datos deben poseer estas cualidades que pongan en evidencia su bondad para recolectar los datos que requiere el investigador. Muchas veces, el investigador fracasa en sus esfuerzos al recolectar los datos debido a que sus instrumentos no poseen las cualidades mínimas y le proporcionan datos falsos o equivocados. Un buen instrumento debe reunir ciertas cualidades que lo tipifiquen como tal.

Son muchas las cualidades que deben poseer los instrumentos de acopio de datos, pero las más importantes son las siguientes: validez, confiabilidad, objetividad, amplitud, practicabilidad y adecuación.

7.1 Validez

La validez es una cualidad que consiste en que las pruebas midan lo que pretenden medir. Las pruebas deben medir las características específicas de las variables para las cuales fueron diseñadas. Las pruebas que no poseen validez no tienen utilidad alguna. La validez también se denomina veracidad, exactitud, autenticidad, o solidez de la prueba.

La validez se refiere a los resultados de la prueba, no a la prueba misma. Estos resultados no se expresan en términos categóricos: resultados válidos o resultados no válidos, sino que estos resultados se expresan en forma de una continuidad o progresión, así los resultados serán de escasa validez o de mucha validez, pasando por múltiples situaciones intermedias.

La validez de los resultados de una prueba sólo tiene sentido dentro del contexto en el que ocurre la prueba. Por ejemplo, un ítem de resolución de problemas de sumas será válido si lo que se desea es medir esta habilidad en los estudiantes. Pero este mismo ítem no será válido si lo que se desea es explorar las habilidades de multiplicación que habrían desarrollado.

Sin embargo, las pruebas no poseen validez universal. Una prueba válida para una situación determinada puede carecer de validez para otra.

La validez puede ser de varios tipos: validez de contenido, validez de construcción, validez predictiva, validez concurrente y validez estadística.

Validez de contenido

Denominada también validez lógica o de "muestreo". Consiste en que los contenidos o conceptos planteados en los ítemes correspondan con los previstos en los objetivos del aprendizaje. La validez de contenido adquiere mayor importancia cuando se trata de comprobar resultados del aprendizaje.

En otras palabras, la validez de contenido es el grado de fidelidad con el que una prueba refleja el universo de reactivos del cual se extrajeron los ítemes. 'Universo de reactivos' es un concepto teórico que alude a todos los posibles ítemes que se podrían formular sobre un determinado tema. Por ejemplo, si se desea explorar cuánto saben los estudiantes de medicina acerca de la Histología, se tendría que redactar todos los ítemes posibles sobre este tema, lo que resulta un imposible. Sin embargo, teniendo en cuenta que éste es un concepto teórico, los ítemes que se

redacten de una prueba de Histología se considerarán una muestra de este universo de reactivos y, por tanto, se asume que lo representan fielmente. Así, la prueba de Histología que se prepare, será capaz de medir realmente los conocimientos de Histología y no otro tipo de conocimientos.

Para determinar la validez de contenido se analiza el universo de contenidos respectivo y se selecciona una muestra representativa de ellos, la que debe ser explorada por la prueba. Por ejemplo, si se trata construir una prueba de Álgebra, se debe partir de los objetivos del aprendizaje previstos y examinar los textos que existen sobre la materia con la finalidad de determinar el universo de contenidos y luego decidir qué contenidos debe abarcar la prueba y qué proporción de ítemes debe asignarse a cada uno de los aspectos de la materia que se evalúa.

Si se desea saber si una prueba ya publicada tiene validez de contenido para los estudiantes a quienes se pretende aplicar, se deberá comparar los objetivos de aprendizaje previstos para ellos con el contenido de la prueba a aplicarse.

La validez de contenido se puede comprobar según los siguientes criterios:

a) *Criterio curricular o programático*: Los contenidos a medir deben estar previstos en el currículo o en el programa de enseñanza. Para el caso se puede elaborar una tabla de cotejo que permita determinar que todos los contenidos que aparecen en la prueba han sido considerados en el programa curricular.

b) *Criterio bibliográfico*: Los contenidos a medir deben aparecer en los diferentes textos de la materia.

c) *Criterio de pertinencia*: Los contenidos a medir deben estar adecuados a las capacidades intelectuales de los sujetos que responderán la prueba. De este criterio se desprende la necesidad de determinar los índices de dificultad, tanto de los ítemes como de la prueba.

d) *Criterio de utilidad social*: Los contenidos a medir deben tener significado o importancia para la vida práctica.

La validez de contenido se determina generalmente mediante el juicio de expertos, sin embargo, hay técnicas como la aplicación de pruebas paralelas o el estudio de errores comunes, que proporcionan la información requerida sobre este tipo de validez.

Validez de constructo o de concepto

El constructo es un concepto elaborado al más alto nivel de abstracción y sirve para referirse a una característica o aspecto de la conducta humana, como por ejemplo, la *habilidad mecánica*, la *inteligencia* o el *aprendizaje*.

La validez de constructo es el grado de correspondencia o congruencia que existe entre los resultados de una prueba y los conceptos teóricos en los que se basan los temas que se pretenden medir. La validez de constructo trata de establecer en qué medida la prueba tiene en cuenta los aspectos que se hallan implícitos en la definición teórica del tema a medirse.

Este tipo de validez es el más importante, pues si se elabora una prueba sobre Anatomía, los resultados de la prueba deben medir realmente conocimientos acerca de lo que entendemos por Anatomía y no sobre cualquier otro concepto. En este caso, la dificultad radica en la capacidad que se tenga para definir correctamente el constructo que se desea explorar.

Se denomina también validez de elaboración o de hipótesis de trabajo. Se determina en base al juicio de expertos o a los tipos y grados de relaciones existentes entre el resultado de la prueba y las medidas de otros factores asociados. Es el tipo de validez más importante. Cuando se trata de determinar la validez de constructo, se procura analizar la naturaleza de las características que el investigador se ha propuesto medir.

La validez de constructo no sólo evalúa la prueba, sino también la teoría en la que se apoya la prueba. Cuando los datos obtenidos confirman las predicciones de la teoría, significa que, al menos por el momento, se han hallado evidencias que convalidan la hipótesis e indican que la prueba mide las relaciones previstas por esta última.

Si las predicciones no son confirmadas, se debe suponer que la causa puede ser una de las siguientes: la prueba no mide el concepto, la hipótesis es incorrecta, se requiere un nuevo análisis, o el diseño de la investigación no es el más adecuado.

Validez predictiva

La validez predictiva es la capacidad que tienen las pruebas de predecir acontecimientos futuros, tales como el éxito que un individuo alcanzará en sus estudios o en su trabajo. Para determinar la validez predictiva se procede del siguiente modo:

1. Se administra la prueba.
2. Se espera que se produzcan los desempeños pronosticados.
3. Se compara los puntajes de la prueba con los desempeños reales del sujeto.

Suponiendo que una prueba ha sido elaborada para predecir el éxito que los alumnos lograrán durante el primer año de estudios universitarios, para determinar su validez predictiva, se administrará la prueba a una amplia muestra de alumnos

del 5º año de Secundaria, seleccionados al azar. Una vez que estos sujetos hayan terminado el 1er. año de estudios universitarios, se cotejarán ambas series de puntuaciones. Cuanto más alta sea la correlación entre ambas series de puntuaciones, la prueba tendrá mayor capacidad predictiva.

Sin embargo, a nivel teórico, el pronóstico sólo puede ser válido si la comparación con los desempeños reales se hace en situaciones iguales o semejantes a las existentes al momento de aplicar la prueba, cosa que no habría ocurrido en la situación descrita en el párrafo anterior.

Validez concurrente

La validez concurrente es la correlación que puede hallarse entre las puntuaciones logradas luego de la aplicación de la prueba, con respecto a las puntuaciones obtenidas por los mismos sujetos en otras mediciones realizadas simultáneamente.

El procedimiento para determinar la validez concurrente es similar al que se emplea para verificar la validez predictiva, sólo que la correlación puede hallarse casi de inmediato, ya que los datos se obtienen simultáneamente.

Los puntajes que los estudiantes obtienen en una nueva prueba se pueden comparar con las calificaciones que recibieron últimamente en la materia o con los puntajes que los mismos alumnos alcanzaron en otras pruebas similares, cuya validez ya fue comprobada. Por ejemplo: En lugar de esperar varios años para determinar si una prueba de interés vocacional es capaz de predecir el éxito en determinada profesión, se puede cotejar esta prueba con las pautas de interés de las personas que han alcanzado éxito en la profesión correspondiente.

Validez estadística o empírica

Es el tipo de correlación que existe entre los resultados de la prueba y la situación real del sujeto en los aprendizajes cuyos resultados se mide. Para establecerla se usa las siguientes técnicas:

1. Comparación con las diferentes calificaciones obtenidas.
2. Incremento del porcentaje de éxito por edades y grados sucesivos.
3. Análisis de las diferencias entre los puntajes obtenidos por 2 grupos diferentes de alumnos.

7.2 Confiabilidad

El término confiabilidad proviene de la palabra fiable, y ésta a su vez de fe. La confiabilidad es el proceso de establecer cuan fiable, consistente, coherente o estable es el instrumento que se ha elaborado. Por eso, cuando el investigador, al referirse a una persona dice que ésta es confiable, quiere decir que le inspira confianza, que tiene fe en ella y en cambio cuando se refiere a otra persona puede decir que no le merece confianza, que duda de su conducta futura. Una persona considerada confiable es aquella que actuará en el futuro, del mismo modo como ha actuado en el pasado. Así también, un instrumento de acopio de datos será confiable, inspirará confianza, cuando al ser aplicado en repetidas ocasiones arroje los mismos resultados.

Cuando se realizan procesos de medición, el puntaje observado de una variable equivale al puntaje verdadero más el margen de error. El puntaje observado es el valor que se obtiene al medir una variable, sin embargo, éste no es el puntaje verdadero; el puntaje verdadero es el resultado de la medición exacta de la variable. Por ejemplo, los estudiantes tienen un determinado nivel de éxito académico en sus estudios que es precisamente el que se quiere conocer. Este valor es único y cuando se hacen sucesivas mediciones, se obtienen diversos puntajes aproximados a este valor, pero no se obtiene el verdadero valor de la variable, precisamente porque el investigador no dispone de un instrumento tan perfecto que sea capaz de medir la real magnitud de, en este caso, el éxito académico de los estudiantes, que efectivamente existe pero no lo puede conocer. De esta idea parte aquello de que toda medición debe ser isomórfica con la realidad, es decir, que las mediciones que se obtengan, aunque no sean exactas, deben ser, por lo menos, aproximadas a la real magnitud del fenómeno que se estudia.

La humanidad ha avanzado mucho en la medición del tiempo, entre otros fenómenos que ha medido, y los resultados de tales mediciones son asombrosamente aproximadas a la real magnitud del tiempo, sin embargo, los científicos declaran que sus mediciones del tiempo se están realizando con aproximaciones de décimas, centésimas o milésimas de segundo, pero nunca logran medir el tiempo exactamente y sin márgenes de error. Sin embargo, se puede decir que la medición del tiempo es isomórfica con la realidad, porque el margen de error es muy pequeño, tan pequeño que resulta insignificante.

Pero es el caso que se constata que los instrumentos que se disponen para medir las variables del comportamiento, tales como la *inteligencia*, el *éxito académico*, la *motivación por los estudios*, etc., no son tan precisos como los instrumentos que la humanidad dispone para medir el tiempo o el espacio. Ahí radica, entonces, la importancia de elaborar instrumentos confiables que permitan al investigador aproximarse, lo más cerca posible, al valor real de las variables y

para ello debe el investigador reducir el margen de error con el que hace las mediciones. En este sentido, si teóricamente un estudiante tiene 17,57 puntos de éxito académico en sus estudios y el investigador ha obtenido, con los instrumentos que dispone, una puntuación de 17,13, se podrá decir que ha hecho una medición exacta y el margen de error es relativamente pequeño, pero si le asigna un puntaje de 12,45, entonces el investigador se está alejando del valor verdadero del éxito académico del estudiante. Sin embargo, con una prueba confiable, en las repetidas oportunidades en que se realicen las mediciones, se obtendrán valores muy próximos al valor teórico y, en este caso, se podría considerar que los resultados de estas mediciones son más cercanos al puntaje verdadero, 17,57, son más plausibles que el valor de 12,45 obtenido con una prueba poco confiable.

Como se ha dicho, la confiabilidad es la seguridad, exactitud, precisión o consistencia que debe poseer una prueba. Una prueba es confiable si al aplicarla en reiteradas ocasiones a los mismos sujetos y en idénticas condiciones se obtiene iguales resultados. Por ejemplo, si un estudiante alcanza un puntaje de 110 en una prueba de inteligencia, debe lograr aproximadamente el mismo resultado si, una semana más tarde, se le administra una forma equivalente de la misma la prueba.

La confiabilidad se expresa mediante el índice de confiabilidad. El índice de confiabilidad perfecto es 1, muy difícil de alcanzar. Los índices de confiabilidad aceptables oscilan entre 0,66 y 0,71, como mínimo, es decir, decimales que tienden a acercarse a la unidad.

El índice de confiabilidad de una prueba se obtiene aplicando la siguiente fórmula:

$$C_f = \frac{n}{n-1}\left[1 - \frac{\bar{x}(n-\bar{x})}{n\sigma^2}\right]$$

Donde:

C_f = Coeficiente de confiabilidad
n = Puntaje máximo alcanzado.
$\bar{x}$ = Promedio.
σ = Desviación standard de las puntuaciones de la prueba.

Sean los siguientes datos:
n = 18
$\bar{x}$ = 14,27
σ = 3,1

Entonces:
$$C_f = \frac{18}{18-1}\left[1 - \frac{14,27(18-14,27)}{18(3,1)^2}\right]$$

$$Cf = 1,0588 \text{ x } \left[1 - \frac{53,2271}{172,98}\right]$$

Cf = 1,0588 x [1 – 0,3077]

Cf = 1,0588 x 0,6923

Cf = 0,73

Con los datos proporcionados, el índice de confiabilidad hallado es 0,73.

La tabla de Küder Richardson permite interpretar este valor hallado:

0,53 a menos	=	Confiabilidad nula
0,54 a 0,59	=	Confiabilidad baja
0,60 a 0,65	=	Confiable
0,66 a 0,71	=	Muy confiable
0,72 a 0,99	=	Excelente confiabilidad
1,0	=	Confiabilidad perfecta

Como se ha obtenido el valor de 0,73, se deduce que la supuesta prueba analizada tiene un excelente coeficiente de confiabilidad

Otro ejemplo. Sean los siguientes datos:

n = 98

x = 70

σ = 6,8

Entonces:
$$C_f = \frac{98}{98-1}\left[1 - \frac{70(98-70)}{98(6,8)^2}\right] = 0,57$$

El resultado 0,57 significa que la prueba tiene bajo índice de confiabilidad.

Factores que permiten mejorar la confiabilidad de una prueba

Existen muchos factores que permiten mejorar el índice de confiabilidad. Mucho depende del cuidado y del esmero que ponga el investigador al elaborar la prueba y también de su experiencia en este tipo de tarea. Sin embargo, en forma muy puntual, se puede mencionar los siguientes factores:

a) Adecuado número de ítemes. Una prueba con pocos ítemes no es muy confiable.

b) Homogeneidad de los elementos de la prueba. Cuanto mayor sea la homogeneidad de los elementos de una prueba, la prueba será más confiable.

c) Índice de discriminación de los ítemes. A mayor índice de discriminación de los ítemes, mayor confiabilidad de la prueba.

d) Grado de dificultad de la prueba. Las pruebas que tienen un índice de dificultad media son más confiables.

e) Grado de representatividad de la prueba. La prueba debe contener ítemes que constituyan una muestra representativa de las conductas a medir.

Técnicas para incrementar la confiabilidad de una prueba

Como quiera que la confiabilidad depende de la relación que existe entre el puntaje verdadero y el puntaje observado, es lógico suponer que cuanto menor sea el margen de error, el puntaje observado se hallará más cercano al puntaje verdadero, que es un concepto teórico. De modo que si se desea incrementar la confiabilidad de un instrumento de acopio de datos, lo que se deberá hacer es tratar de disminuir el margen de error. Las siguientes son las técnicas más eficaces para incrementar la confiabilidad.

- *Aumentar el número de ítemes o de observaciones de una prueba*. Cuanto mayor sea la muestra extraída del universo de conductas que se está estudiando, es más probable que tal muestra sea más representativa y más confiable.
- *Eliminar los ítemes difíciles o ambiguos*. Los ítemes difíciles o ambiguos son respondidos de distintas formas, lo que hace disminuir la confiabilidad de la prueba.
- *Estandarizar las condiciones en las que se administra la prueba*. Todas las personas a quienes se aplique la prueba deben resolverla en igualdad de condiciones. No puede ser que mientras un grupo de sujetos resuelva la prueba en un ambiente adecuado y sin ruidos ni interferencias, otro grupo resuelva la prueba en un ambiente inapropiado y con muchas interferencias.
- *Emplear ítemes con adecuado índice de dificultad*. Toda prueba que contenga ítems demasiado fáciles o demasiado difíciles no mide el real desempeño de los examinados.
- *Minimizar los efectos de factores externos*. Es preferible no someter las pruebas a procesos de confiabilidad en momentos en que la sociedad vive circunstancias especiales, como son las épocas de elecciones, la sucesión de

varios días feriados, los días de convulsión social, etc. Es recomendable postergar la prueba para momentos más oportunos debido a que es muy probable que tales eventos influyan en el ánimo de los sujetos y produzcan distorsiones en los resultados de la medición.

- *Estandarizar las instrucciones*. Diferentes investigadores que administran la prueba en distintos grupos, deberán leer las mismas instrucciones y presentar la prueba en las mismas condiciones.
- *Mantener procedimientos de calificación coherentes*. Esto es recomendable especialmente cuando la prueba contiene ítemes de respuesta abierta. En este caso, un evaluador no actúa del mismo modo cuando califica la primera prueba que cuando califica la última de una cantidad grande de pruebas. Es necesario que la actitud del evaluador sea consistente y que disponga de criterios objetivos según los cuales pueda guiarse al momento de asignar puntajes a las respuestas.

Técnicas para establecer la confiabilidad de una prueba

Para determinar la confiabilidad de una prueba, por lo general, se compara dos versiones de la misma prueba. La forma numérica de expresar esta comparación es a través de un coeficiente de correlación (*r*) que puede existir entre ellas. Cuanto más alto sea el coeficiente de correlación positiva entre las dos versiones de la prueba, la prueba será más confiable. Demás está decir que el coeficiente de correlación perfecto es 1,00 y para asignar alto grado de confiabilidad a una prueba se debe esperar obtener coeficientes de, más o menos, 0,80.

Las principales técnicas para determinar la confiabilidad son las siguientes:

a. *Técnica de la replicación* o *test - retest*: La prueba o test se aplica dos o más veces al mismo grupo de estudiantes y se correlacionan las series de puntuaciones obtenidas. Para que la prueba sea confiable se espera que la correlación sea alta. La desventaja de este método consiste en que los resultados del post test puedan ser influenciados por el pre test, habida cuenta que el test que se aplica es el mismo, así como el grupo que responde la prueba.

b. *Técnica de la aplicación de dos versiones paralelas de una misma prueba.* Se elabora dos versiones paralelas de la misma prueba y se aplica, cada una de ellas, a dos grupos distintos. En base a las dos series de puntuaciones así obtenidas se calcula el coeficiente de correlación de tales series de puntuaciones; si la correlación es alta, se puede establecer que la prueba es confiable. Esta técnica se aplica cuando se sospecha que el recuerdo o la práctica en el pre test, pueda influir en los resultados del post test. La desventaja de esta técnica es la dificultad

para elaborar versiones paralelas de la misma prueba y cuidar que los grupos a los que se aplica sean relativamente iguales entre sí.

c) *Técnica de aplicación de la prueba en mitades*. Esta técnica consiste en dividir la prueba, al azar, en mitades y aplicarla al mismo grupo en un solo momento. El coeficiente de correlación se obtiene correlacionando la serie de puntuaciones obtenidas en la primera mitad de la prueba con las puntuaciones obtenidas en la segunda mitad. Si por ejemplo una prueba tiene 40 ítemes, se considera la primera mitad a los 20 primeros ítemes y la segunda mitad serán los 20 ítemes restantes. Otra técnica más efectiva es considerar primera mitad a los ítemes impares y segunda mitad a los ítemes pares. Para dividir la prueba en dos mitades empleando este criterio, se diseña una hoja de respuestas de modo tal que las respuestas a los ítemes impares se coloquen en la columna de la derecha y las respuestas a los ítemes pares se coloquen en la columna de la izquierda. De este modo es posible obtener dos series de puntuaciones en un mismo tiempo y con los mismos sujetos, lo que neutraliza la desventaja de la administración de test, es decir, la influencia que puede producir el hecho de aplicar dos veces una prueba a un mismo sujeto. El siguiente gráfico ilustra lo dicho:

HOJA DE RESPUESTAS

1	a	b	c	d
3	a	b	c	d
5	a	b	c	d
7	a	b	c	d
9	a	b	c	d
11	a	b	c	d
13	a	b	c	d
15	a	b	c	d
17	a	b	c	d
19	a	b	c	d

2	a	b	c	d
4	a	b	c	d
6	a	b	c	d
8	a	b	c	d
10	a	b	c	d
12	a	b	c	d
14	a	b	c	d
16	a	b	c	d
18	a	b	c	d
20	a	b	c	d

Gráfico N° 3

Con este método, la prueba se aplica una sola vez, se evita los efectos del pre test y se puede establecer que la prueba ha sido aplicada a dos grupos iguales de sujetos.

a. *Técnica de la administración en tiempo diferido*. Una misma prueba debe ser administrada al mismo grupo en dos ocasiones diferentes. De este modo se obtiene las dos series de puntuaciones necesarias para calcular el coeficiente de correlación. Es recomendable que el lapso entre la primera y la segunda aplicación no sea ser muy breve, para evitar la llamada *administración de test*,

es decir, la influencia de la primera aplicación en los puntajes que se obtengan en la segunda aplicación.

7.3 Relaciones entre la confiabilidad y la validez

La confiabilidad y la validez son cualidades estrechamente relacionadas entre sí. Sin embargo, una prueba puede ser confiable pero no necesariamente válida, por eso, una prueba no puede ser válida, si previamente no se ha establecido su confiabilidad. Esto significa que la confiabilidad es una condición necesaria pero no suficiente para establecer la validez de una prueba. Por ejemplo, una prueba para medir *aprendizajes matemáticos* puede ser altamente confiable, pero si pretendemos que esa prueba mida *comprensión lectora*, obviamente esta prueba ya no será válida, aunque siga siendo confiable. En resumen, una prueba puede ser confiable y evaluar de manera consistente un constructo, pero si este constructo no es el mismo que se está estudiando, la prueba no será válida.

Como se puede notar, en las ciencias sociales aún existen serias dificultades para medir las variables que son materia de investigación. En cambio, en las ciencias naturales, existe un mayor nivel de desarrollo al respecto, debido principalmente a que la mayoría de las variables científico naturales están plenamente operacionalizadas y existen diversos instrumentos para medirlas. Las variables del comportamiento humano, por ser elusivas e impredecibles, ofrecen mayores dificultades, tanto para su operacionalización como para su medición. Por eso es que en las ciencias sociales no se dispone de muchos instrumentos para medir estas variables. De ahí que resulta de enorme responsabilidad del científico social realizar los mayores esfuerzos para construir instrumentos válidos y confiables que le permitan conocer, con aproximaciones muy cercanas a lo real, las características de los fenómenos que estudia. De lo contrario, si no se miden correctamente las variables científico sociales, se corre el riesgo de arribar a conclusiones equivocadas en la investigación y de nada habría valido haber planteado buenas hipótesis o haber elegido las mejores estrategias para contrastarlas. El éxito en las investigaciones científico sociales depende, muy significativamente, de la calidad de los instrumentos con los que se estudian los fenómenos.

7.4 Objetividad

La objetividad consiste en que la prueba debe producir los mismos puntajes sea quien fuere el que evalúe las respuestas. Para lograr esta condición, las pruebas deben disponer de normas de asignación de puntajes o claves de respuestas para no tener la necesidad de recurrir a juicios subjetivos para decidir si cada respuesta es correcta o incorrecta.

7.5 Amplitud

Es la adecuada extensión que debe tener la prueba. Una buena prueba debe explorar, en un tiempo determinado, la mayor cantidad de información. Las pruebas no deben ser muy amplias ni muy escuetas, es recomendable que las pruebas sean de amplitud mediana.

7.6 Practicabilidad

Una prueba debe ser práctica o aplicable. Para ello debe estar diseñada de tal modo que su administración e interpretación sea fácil y de bajo costo.

También es conveniente tener en cuenta su utilidad social. La prueba debe conducir a soluciones prácticas, que ofrezcan alguna utilidad en la solución de los problemas de la vida diaria.

7.7 Adecuación

Cuando se diseña una prueba, es necesario determinar si ésta es apropiada para el fin que se persigue. Habrá que tener en cuenta si la prueba proporciona los datos necesarios, si permitirá obtener las medidas en el grado de precisión que el investigador desea, si será apropiada para la edad y características de los sujetos, para el momento y la localidad en que se pretende administrarla, o cuándo preferir entre dos pruebas igualmente confiables y válidas. Por estas consideraciones, resultan más adecuadas las pruebas que tienen formas paralelas y normas específicas para su empleo o aplicación.

8. PRINCIPIOS PARA CONSTRUIR PRUEBAS

Para construir una prueba, se deben tener en cuenta los siguientes principios:

a. Toda prueba debe cumplir un propósito bien definido. Los propósitos pueden ser el diagnóstico, el pronóstico, la investigación, etc. Según sus propósitos, las pruebas adquieren características particulares en relación con el índice de dificultad de los ítems, la interpretación de los resultados, el momento en que deben ser aplicados, etc.

b. Las pruebas deben recoger evidencias sobre determinados productos del aprendizaje en función de objetivos previamente determinados. Cada ítem debe medir algún conocimiento o habilidad específica. El tipo de ítem dependerá del tipo de aprendizaje que se desee evaluar. Por ejemplo, si se pretende medir conocimientos específicos, no es lo mismo utilizar una pregunta de ensayo que un ítem de verdadero – falso.

c. Las pruebas se construyen en base a un muestreo del universo de contenidos a evaluar. El diseño previo de la estructura de la prueba asegura un buen muestreo.

d. Según sus propósitos, las pruebas deben tener un nivel apropiado de dificultad. Sin embargo, en las pruebas de diagnóstico, o en las formativas, lo que interesa es la respuesta personal de los estudiantes para analizar sus deficiencias de aprendizaje y recomendar las alternativas que les permitan lograr los objetivos.

e. Los ítems de las pruebas deben estar elaborados de modo tal que en la respuesta incluyan sólo el logro de un determinado aprendizaje. Lo que importa es que un ítem mida determinados productos de aprendizaje y no habilidades pre requeridas para ello.

f. Los ítems se deben elaborar de tal forma que sólo cuando el alumno haya logrado el producto del aprendizaje, sea capaz de formular la respuesta correcta.

9. TÉCNICA PARA ELABORAR PRUEBAS

Como se requiere mayor cuidado para elaborar tests o pruebas que midan los resultados del aprendizaje, es necesario aplicar una técnica específica para construirlos. Si, por ejemplo, se desea elaborar una prueba para medir el aprendizaje de estudiantes universitarios en un curso de Física, se procede de la siguiente manera:

1. ***Determinación de los propósitos*** y los usos para los cuales servirá la prueba. Los propósitos para los cuales se elaboran las pruebas son múltiples, pero los más conocidos son los siguientes: motivación, orientación, diagnóstico, pronóstico, administración, investigación, validación, o evaluación de los resultados del proceso educativo. En este momento se debe establecer si la prueba que se va a elaborar servirá para diagnosticar, pronosticar o simplemente para validar el instrumento. En este caso se desea elaborar una prueba para diagnosticar el nivel de aprendizaje en el curso de Física.
2. ***Identificación del objeto a evaluar***. Una vez establecido el objetivo, se *determina el objeto a evaluar*, es decir la identificación de los sujetos a quienes se les aplicará la prueba. En este caso, los estudiantes del III Ciclo de la carrera profesional de Ingeniería Mecánica. Cuando la prueba se elabora con fines de investigación, los sujetos a quienes está dirigida la prueba constituyen la población y los sujetos a quienes se les aplica, deben constituir una muestra representativa de la población que se estudia.
3. ***Elaboración de la estructura de la prueba***. La estructura de la prueba se grafica mediante un cuadro de doble entrada en el que se consignan los datos requeridos. Este cuadro también permite identificar los sub tests que contiene la prueba. En el ejemplo propuesto, se trata de medir el aprendizaje en función

de dos variables: a) los objetivos clasificados según una Taxonomía, en este caso la de Bloom, y b) los contenidos ordenados en Unidades de Aprendizaje. Luego se identificarán los aspectos a medir, que resultan del cruce de estas dos variables. Se decidirá si se van a medir todos los aspectos que aparecen en el Cuadro Nº 2 o solamente algunos de ellos, según su importancia.

En el cuadro de doble entrada según el cual se representa la estructura de la prueba se consignan los siguientes datos:

a. Los objetivos o capacidades que se pretenden medir deben estar identificados en forma clara y precisa en función de una determinada taxonomía de objetivos. En este caso, como se ha dicho, se usará la Taxonomía propuesta por Bloom que considera tres dominios del aprendizaje: el cognitivo, el afectivo y el motriz. En el presente ejemplo se han considerado los tres primeros niveles del dominio cognitivo. Cada objetivo debe estar en correspondencia con un determinado nivel del dominio de aprendizaje. En este caso se medirán los resultados del aprendizaje en función de dos variables: los objetivos clasificados en la Taxonomía de Bloom y los contenidos, ordenados en Unidades de Aprendizaje.

b. Los contenidos en función de los objetivos o capacidades.

c. La extensión de la prueba, que depende de los siguientes factores:

- Propósitos de la prueba.
- Importancia del objetivo.
- Grado de confiabilidad que se desea asignar.
- Tipo de ítems que se emplearán.
- Tamaño y complejidad de la pregunta.
- Modo y tiempo probable de aplicación.
- Puntaje total de la prueba.
- Escala en la que se calificará.
- Edad de los sujetos.
- Nivel de capacidad intelectual.

ESTRUCTURA DE UNA PRUEBA

Obj. / Cont.	Información	Comprensión	Aplicación	Total	Puntaje (total x peso)
U. D. Nº 1	5	5	--	10	15
U. D. Nº 2	4	3	2	9	16
U. D. Nº 3	2	4	3	9	19
Total	11	12	5	**28**	
Peso	1	2	3		
Puntaje (total x peso)	11	24	15		**50**

Cuadro N° 2

En el ejemplo propuesto se observa que se elaborará una prueba con 28 ítems que darán un puntaje de 50 puntos que resultan de multiplicar el total de ítems de cada nivel de dominio por el peso asignado.

La asignación de pesos depende de la importancia que asigne el investigador a un determinado nivel o dominio de aprendizaje. En el presente caso se observa que se está asignando mayor peso (3) a la aplicación de la información.

4. ***Selección de los ítems***. Este paso consiste en seleccionar los ítems, reactivos o preguntas que debe contener la prueba. En este momento también se decide acerca del tipo de ítem a utilizar en función del nivel y tipo de resultado de aprendizaje o capacidad que se desea medir, de las ventajas y desventajas de los diferentes tipos de ítems, del modo en que se utilizarán estos, de las características de los estudiantes, del tiempo que se requerirá para construirlos, del modo de administrar el test, de la escala de medición a emplear, y de otros factores que considere el investigador.

   El mejor ítem es aquel que mide, del modo más directo posible, el resultado de aprendizaje deseado. También se debe tener en cuenta si los ítems poseen adecuados índices de dificultad y de discriminación.

5. ***Elaboración de los ítems***. Si el investigador dispone de un banco de ítems, reactivos o preguntas, seleccionará los ítems más adecuados para la prueba que está construyendo, teniendo en cuenta las necesidades de la investigación. En caso contrario, deberá elaborar los ítems previstos en la estructura.

6. ***Elección de la escala de medición***. En este punto es necesario establecer si la escala será decimal, vigesimal, centesimal o cualquier otra.

7. ***Asignación de los puntajes de la prueba***. Los puntajes de la prueba deben estar de acuerdo con los objetivos, contenidos y o capacidades que se pretende medir o evaluar. Según la extensión de la prueba, se decidirá con respecto a su división en partes menores llamadas sub tests. Aquí se realiza el ensamblaje de los ítems, que consiste en el ordenamiento de los ítems agrupados para cada uno de los sub tests de la prueba.

8. ***Asignación de pesos***. Se asignan pesos a cada uno de los sub tests de la prueba. De preferencia los pesos deben asignarse porcentualmente en función de la importancia de los temas tratados. En el presente ejemplo, el nivel de *información* tiene peso 1, el nivel de *comprensión* tiene peso 2 y el nivel de *aplicación* tiene peso 3. Así, la prueba que se va a construir tendrá 28 ítems que multiplicados por sus respectivos pesos, producirán 50 puntos.

9. ***Jerarquización lógica de los ítems***. Es conveniente presentar previamente los ítems simples y los más complejos después. También conviene agrupar los ítems por afinidad temática u otros criterios.

10. ***Elaboración de la Tabla de Especificaciones de la prueba***. La Tabla de Especificaciones es el conjunto de instrucciones que deben observarse para usarla convenientemente. En este punto se decidirá si la aplicación de la prueba debe ser individual o grupal, si se aplicará a los estudiantes de un determinado ciclo de su formación profesional, etc.

    Para hacer un mejor uso de las pruebas, se debe respetar sus especificaciones técnicas. Las pruebas que cuentan con especificaciones técnicas pueden ser generalizadas para otras poblaciones. Las especificaciones técnicas, redactadas de manera clara y concisa, deben referirse a los siguientes aspectos:

a) Grado de dificultad que ofrece la prueba. Cuando la medición se realiza en función de una norma, como el caso de los procesos de selección de postulantes, los ítems deberán tener un 50% de dificultad.

b) Cuando la medición se realiza en base a un criterio, el nivel de dificultad variará de acuerdo con el nivel de rendimiento de los estudiantes y la variabilidad que exista entre ellos. La dificultad del ítem se determinará, sólo en parte, por el nivel de rendimiento que ha alcanzado el estudiante.

c) Forma de presentación de la prueba. Se debe diseñar un formato que facilite la lectura de las preguntas y la tabulación de los resultados. Las pruebas generalmente se presentan en forma de cuadernillo.

d. Forma cómo deben responder los sujetos. Las respuestas pueden ser orales, escritas o manifestadas a través de acciones concretas. Las pruebas objetivas son más fáciles de revisar y se responden en hojas separadas. Las de ensayo se contestan normalmente por escrito. La respuesta oral es más fácil y rápida de evaluar que una escrita o de tipo ensayo.

e. Tiempo que demande la aplicación del test.

f. Normas que ayuden a quienes los administraran para establecer si los sujetos alcanzaron el nivel de habilidad promedio, lo superaron o si se hallan por debajo de él.

   Para formular las normas, el diseñador de la prueba extrae una muestra de la población a la cual se destina la prueba, administra la forma final a los sujetos seleccionados y elabora las reglas de acuerdo con los datos obtenidos.

11. ***Realización del Estudio Piloto***. Antes de aplicar una prueba, es recomendable realizar un estudio piloto. Cuando los sujetos del estudio piloto terminan de responder el test, el investigador analiza sus respuestas y revisa las instrucciones que no fueron comprendidas claramente. También debe superar los inconvenientes relacionados con el formato elegido y elimina o mejora los ítems que mostraron poseer escasa utilidad, etc. Una vez realizadas las correcciones, efectúa un nuevo examen para asegurarse que todos los aspectos de la habilidad que se desea medir se hallan representados en el test, con la debida proporción.

10. INDICADORES DE LA CALIDAD DE UNA PRUEBA

La calidad de una prueba se puede establecer empleando los siguientes criterios:

a. Observaciones más frecuentes de los estudiantes con respecto:
   - Al excesivo número de preguntas referidas a uno o más temas y muy pocas al resto de los contenidos explorados.
   - A la redacción ambigua de las pruebas que hasta los alumnos más aplicados no pueden entenderlas.

b. Demasiadas preguntas de los estudiantes durante el examen.

c. Mayor tiempo del previsto que necesitaron los estudiantes para responder a las preguntas.

d. Situación en la que los estudiantes más aplicados no respondieron satisfactoriamente la prueba.

e. Análisis de los temas para determinar su calidad. Para ello se debe obtener el grado de dificultad y el índice de discriminación de cada ítem.

11. GRADO DE DIFICULTAD DE LA PRUEBA

El grado de dificultad de la prueba se determina aplicando la siguiente fórmula:

$$Gd = \frac{\bar{x}}{Pm} \times 100$$

Donde:

Gd = Grado de dificultad de la prueba.

$\bar{x}$ = Promedio de los puntajes obtenidos.

Pm = Puntaje máximo posible de alcanzarse en la prueba.

Ejemplo: Hallar el grado de dificultad de una prueba en la que el grupo ha obtenido un promedio de 13,27 y el puntaje máximo es 20.

Entonces:

$$GD = \frac{13,27}{20} \times 100 = 66\%$$

Para interpretar esta cifra se recurre a la siguiente escala de Kuder-Richardson

81% a más	=	Muy fácil
61% a 80%	=	Relativamente fácil
51% a 60%	=	Dificultad adecuada
31% a 50%	=	Relativamente difícil
11% a 30%	=	Difícil
Debajo del 10%	=	Muy difícil

El valor obtenido, 66, revela que la prueba es relativamente fácil.

Otro ejemplo: Hallar el grado de dificultad de una prueba cuyo promedio es 18,37 y el puntaje máximo es 40.

$$GD = \frac{18,37}{40} \times 100 = 46\%$$

Esta es una prueba relativamente difícil.

12. ÍNDICE DE DISCRIMINACIÓN DE UNA PRUEBA

Para hallar el índice de discriminación de una prueba se forman dos grupos de estudiantes en función de sus respuestas correctas. El grupo superior está constituido por los estudiantes que respondieron mejor y el grupo inferior por los que alcanzaron puntajes más bajos. El índice de discriminación se establece analizando las diferencias de las respuestas correctas del grupo superior y las respuestas correctas del grupo inferior. Se supone que los estudiantes del grupo superior tendrán mejores respuestas que los del grupo inferior y el índice de discriminación será más alto cuanto mayor sea la diferencia de las puntuaciones correctas del grupo superior con respecto a las del grupo inferior.

Para hallar el índice de dificultad de una prueba se aplica la siguiente fórmula:

$$I_d = \frac{pms - pmi}{PM} \times 100$$

Donde:

I_d = Índice de dificultad de la prueba.

pms = Puntaje máximo de respuestas correctas del grupo superior.

pmi = Puntaje máximo de respuestas correctas del grupo inferior.

PM = Puntaje máximo de la prueba.

Si se ha aplicado una prueba a un grupo de 50 estudiantes y se ha elaborado el orden de mérito, de mayor a menor, se considerará grupo superior a los 25 estudiantes que obtienen las más altas puntuaciones y grupo inferior a los 25 restantes que tienen más bajas puntuaciones.

Ejemplo: Hallar el índice de discriminación de una prueba en la que el puntaje máximo de respuestas del grupo superior es 18 y el puntaje máximo de respuestas correctas del grupo inferior es 11. El puntaje total de la prueba es 20.

Entonces:

$$Id = \frac{18 - 11}{20} \times 100 = 35$$

Para interpretar esta cifra se recurre a la siguiente tabla:

40% a más = Buen índice de discriminación.

30% al 39% = Razonable índice de discriminación.

20% al 29% = Regular índice de discriminación.

menos de 19% = Deficiente índice de discriminación.

La prueba analizada posee un razonable índice de discriminación.

Otro ejemplo: Hallar el índice de discriminación de una prueba en la que el puntaje máximo del grupo superior es 15 y el puntaje máximo del grupo inferior es 6. El puntaje máximo de la prueba es 20.

$$Id = \frac{15 - 0}{20}$$

45 % significa que esta prueba posee un buen índice de discriminación.

13. LOS ÍTEMES

Los ítems, llamados también reactivos, son preguntas que se formulan en función de los propósitos para los que ha sido elaborada la prueba. La respuesta a los ítems proporciona la información requerida por lo que de su construcción depende la mayor y mejor información que se pueda obtener.

Para construir un buen ítem se debe tener un profundo conocimiento de la materia que se explora, asumir un conjunto de valores educativos, conocer y entender a las personas que van a responder la prueba y ostentar ciertas capacidades específicas de comunicación verbal.

Un ítem debe estar diseñado de tal manera que motive a la persona a proporcionar un determinado tipo de respuesta. Es muy importante que el ítem esté bien construido y que mida, tan directamente como sea posible, el resultado del aprendizaje que se desea conocer.

14. CLASIFICACIÓN DE LOS ÍTEMES

Existe una amplia variedad de ítemes, cada uno de los cuales posee ventajas y desventajas. Según el tipo de respuesta que se espera de los ítemes, éstos pueden clasificarse en dos grandes grupos: ítemes de respuesta abierta e ítemes de respuesta cerrada.

14.1 Ítemes de respuesta abierta

Los ítemes de respuesta abierta se conocen también como ítemes de tipo cuestionario o ensayo. Como su nombre lo indica, las respuestas las proporcionan los estudiantes quienes las redactan tratando de demostrar los conocimientos que poseen. Este tipo de ítem permite explorar una amplia gama de conocimientos que poseen, sin embargo, su desventaja consiste en que ofrecen muchas dificultades para calificarlos pues los criterios que emplea el evaluador casi siempre son subjetivos.

Los ítems de respuesta abierta todavía pueden variar en dos tipos:

a) los que se aplican con límite de tiempo y,

b) los que se aplican sin límite de tiempo.

En este último caso, los ítems o reactivos sirven para medir la performance de quien responde, es decir, hasta donde es capaz el estudiante de responder las preguntas sin la presión del tiempo.

14.2 Ítems de respuesta cerrada

Como su nombre lo indica, presentan respuestas preestablecidas y la tarea consiste en seleccionar o identificar una de las respuestas planteadas. Este tipo de ítem tiene la ventaja de estar referido a criterios objetivos para su calificación; sin embargo, limitan, por la respuesta ya planteada, otras respuestas que pudiera formular el sujeto.

Los ítems de respuesta cerrada se subdividen en los siguientes tipos:

a) de pareamiento,

b) de respuesta corta,

c) de oraciones incompletas,

d) verdadero-falso,

e) de alternativa múltiple,

f) de plan de redacción,

g) de eliminación de oraciones.

Ítems de pareamiento

Los ítems de pareamiento presentan dos columnas de palabras, frases o símbolos. Las palabras, frases o símbolos de una columna deben asociarse con las de la otra columna tratando de establecer entre ellas correspondencias lógicas o verdaderas. La primera columna es la de las premisas y la segunda columna es la de las respuestas. En esta columna se incluyen las asociaciones correctas y los distractores. La persona que responde el ítem deberá hacer las asociaciones planteadas y para evitar que en el último par de datos responda sin posibilidad de elección es recomendable que la columna de las respuestas contenga más distractores para mantener las posibilidades de elección.

La instrucción que se da a la persona que responderá el ítem es la siguiente: *Relacione las afirmaciones de la columna A con las palabras de la columna B, escribiendo el número que corresponde dentro del paréntesis:*

En este caso, se presenta dos series de datos en dos columnas, por ejemplo, los nombres de pensadores y las corrientes de pensamiento que generaron. El examinado debe relacionar los datos de las columnas escribiendo, dentro del paréntesis, la letra o el número que corresponde a la columna de las preguntas.

Ejemplo:

a. Jean Piaget	conductismo	( )
b. Lev Vigotsky	epistemología genética	( )
c. Skinner	pragmatismo	( )
d. John Dewey	cultural historicismo	( )
	naturalismo	( )

Por lo general, este tipo de ítem se emplea para medir información sencilla mediante la realización de asociaciones. Las relaciones que pueden plantearse son muy diversas, entre ellas las siguientes:

Hombres / realizaciones.
Autores / títulos de obras.
Fechas / hechos históricos.
Términos / definiciones.
Símbolos / conceptos.
Causas / efectos.
Reglas o principios / ejemplo o aplicación.
Plantas o animales / clasificación.
Órganos / funciones.
Objetos / nombres de los objetos.
Principios / ilustraciones.
Máquinas / usos.

Este tipo de pregunta se puede usar también con materiales gráficos, planteando relaciones entre cuadros y palabras, o identificando posiciones en mapas, gráficos y diagramas.

Ítems de respuesta corta

Los ítems de respuesta corta están constituidos por una pregunta que el estudiante debe responder en forma precisa y breve. Ejemplos:

¿Quién sostiene la tesis del aprendizaje significativo?

Respuesta: Ausubel

¿Quién postula una pedagogía naturalista?

Respuesta: Rousseau

Ítems de oraciones incompletas

Los ítems de oraciones incompletas son oraciones en las que se omiten una o algunas palabras, las que se colocan debajo del ítem, junto con otras llamadas distractores. El examinando debe escoger la palabra correcta para que cuando sea colocada en el espacio en blanco, la oración adquiera el sentido deseado. Ejemplos:

Los conceptos que el investigador elabora para referirse a los fenómenos que estudia se denominan .....

a) variables

b) distractores

c) hipótesis

d) problemas

Una estrategia para probar hipótesis consiste en ..... los efectos de las variables independientes, ... los efectos de las variables intervinientes y ..... los posibles errores al medir o recolectar datos.

a) minimizar	*maximizar*	*controlar*
b) cancelar	*operacionalizar*	*analizar*
c) identificar	*controlar*	*predecir*
d) generalizar	*evitar*	*evitar*

Ítems verdadero-falso

Los ítems verdadero o falso plantean una afirmación ante la cual el examinando tiene dos posibilidades de respuesta: verdadero o falso, correcto o incorrecto. Para responderlos, debe encerrar en un círculo la alternativa que considera conveniente.

Para construir ítems del tipo verdadero o falso, se debe seguir las siguientes recomendaciones:

a. Solicitar que corrija los enunciados falsos que aparecen en la pregunta. Un estudiante que identifique correctamente los enunciados falsos puede ser que atine, por casualidad, pero también puede haber elegido la respuesta en base a una mala información.

b. La pregunta de doble alternativa debe estar basada en un material de estímulo específico que haya sido proporcionado previamente. Este tipo de ítem es efectivo cuando se sustenta en algún material que ha sido entregado al examinando, tales como mapas, gráficos, tablas, lecturas, etc. Aquí se proporciona un marco de referencia para responder la pregunta.

c. Se puede usar este tipo de ítem para establecer relaciones de causa - efecto. Para ello se presenta dos afirmaciones unidas por la relación mencionada.

Ejemplos:

Diga si la siguiente relación es verdadera o falsa:

Las variables independientes son las supuestas causas y las variables dependientes son los posibles efectos. *(V) (F)*

Ítems de alternativa múltiple

Los ítems de alternativa múltiple están constituidos por un enunciado o base, que en realidad es la pregunta planteada como un problema o afirmación incompleta y por 4 ó 5 alternativas de respuesta o distractores, uno de los cuales es la respuesta correcta. Al examinado se le solicita que seleccione la respuesta correcta o la mejor respuesta. La alternativa correcta recibe el nombre de clave.

Para construir este tipo de ítem se debe tener en cuenta las siguientes recomendaciones:

1. Que todas las alternativas sean plausibles o verosímiles, es decir todas deben aparecer como verdaderas para que actúen, precisamente, como distractores.
2. Todos los distractores deben contener, en lo posible, el mismo número de palabras.
3. Se debe evitar el uso de distractores en negativo, como por ejemplo, "*no es un factor de* .....". Los enunciados negativos plantean problemas de lectura y proporcionan muy poca información acerca de los conocimientos que posee el sujeto.
4. Si se trata de medir conocimientos de excepciones a reglas o la capacidad de advertir errores, es conveniente subrayarlos o escribirlos con otro tipo de letra.
5. Para economizar tiempo en la lectura, las alternativas deben ser, en lo posible, muy breves.
6. Si algunas palabras se repiten en todas las alternativas o en la mayoría de ellas, se debe redactar nuevamente el enunciado incluyendo en el enunciado las palabras que se repiten, para no tener que presentarlas nuevamente en las alternativas.

7. El problema o enunciado debe plantearse con toda claridad.
8. Se debe evitar el uso de distractores tales como: "*todas las anteriores*", "*todas*", "*ninguna de los anteriores*", "*sólo a y b*", etc.
9. El enunciado debe plantear el problema en forma muy clara para facilitar la comprensión del examinado, antes de que lea las alternativas.
10. El enunciado debe contener únicamente lo necesario para dar al problema o pregunta un carácter claro y específico. Los enunciados extensos, con ideas que nada tienen que ver con la pregunta, reducen la eficacia y la validez de la pregunta. El material que innecesariamente se añade aumenta el tiempo para la lectura del ítem y disminuye el tiempo para la respuesta.
11. Al redactar enunciados orientados a la evaluación de la comprensión o la capacidad de aplicación de principios, se debe utilizar un material novedoso. La mayoría de los ítems asigna demasiado énfasis al aprendizaje memorístico y se descuida otras capacidades más complejas.
12. Se debe asegurar que ante el problema planteado exista sólo una solución o una sola respuesta correcta, o una mejor respuesta.
13. Se debe tener cuidado de no proporcionar indicios o dejar traslucir, involuntariamente, la respuesta correcta. Tales indicios pueden ser incongruencias gramaticales, distractores más extensos, uso de determinadores específicos, etc.

Los ítems de alternativa múltiple presentan las siguientes ventajas:

a. Son más adecuados para medir niveles de comprensión y aplicación del dominio cognoscitivo.
b. Plantean al sujeto la elección de una respuesta entre varias alternativas con lo que se reduce la posibilidad de acierto por azar. Sin embargo las opciones erróneas deben servir para atraer la atención de los examinandos que carecen de información o la poseen en medida muy limitada; por eso los distractores deben ser lógicamente congruentes con el enunciado.
c. Pueden medir varios tipos de aprendizaje, así como productos complejos de procesos de aprendizaje.
d. Ofrecen mayor confiabilidad, ya que el número de alternativas reduce la influencia del azar.
e. El adecuado número y calidad de las alternativas o distractores permite el aprovechamiento de la información para efectos de la evaluación formativa. Si una alternativa errónea es elegida por muchos sujetos, pone en evidencia el error y facilita la detección de problemas de aprendizaje.

Sin embargo, este tipo de ítem presenta las siguientes desventajas:

a. Se limitan a medir aprendizajes expresados en forma de productos verbales, pero no predice la forma en que actuará o qué actitud adoptará el examinando ante lo que sabe.

b. Es difícil obtener suficientes distractores adecuados.

c. No permiten medir pensamientos creativos, organización de ideas, habilidades para organizar pensamientos y construir respuestas.

d. No se adapta bien para medir ciertas habilidades para resolver problemas matemáticos.

Variaciones de los ítems de alternativa múltiple

Este tipo de ítem tiene diversas variantes, algunas de las cuales son las siguientes:

Ítems de respuesta correcta. En este caso, todos los distractores son falsos, menos uno, que es el correcto.

Ejemplo:

Cuando se operacionaliza variables, el lenguaje se usa en la función

a) denotativa
b) connotativa
c) estética
d) metaligüística

Ítems de mejor respuesta. En este caso, todos los distractores son correctos pero sólo uno constituye la mejor respuesta.

Este último tipo de ítem es el más conveniente porque permite explorar otros niveles de aprendizaje como son la comprensión, aplicación, etc., pero su construcción demanda mayores dificultades y entrenamiento de parte del investigador.

Ejemplo:

La investigación científica es, ante todo:

a) un proceso de producción de conocimientos.
b) una actividad eminentemente racional.
c) un procedimiento de comprobación de hipótesis.
d) la secuencia metodológica hipotético deductiva.

15. GRADO DE DIFICULTAD DE UN ÍTEM

El grado de dificultad de un ítem se calcula estableciendo el porcentaje de examinados que respondieron correctamente un determinado ítem. Si un ítem fue respondido por casi la totalidad de los examinados, por ejemplo el 98%, será un ítem muy fácil. Si en cambio fue respondido sólo por el 5%, será muy difícil. Cuanto más difícil sea el ítem más alto será su índice de dificultad.

Es recomendable que los ítems de una prueba posean un grado de dificultad intermedia. Se considera que un ítem tiene un adecuado índice de dificultad si es contestado correctamente por el 60 al 70% de los examinados.

Para hallar el grado de dificultad de un ítem se aplica la siguiente fórmula:

$$Gd = \frac{\sum Rc}{N} \times 100$$

en donde:

Gd = Grado de dificultad

$\sum$ = Sumatoria

Rc = Respuestas correctas

N = Número de alumnos examinados

Ejemplo: Sea una prueba de 20 ítems respondida por 40 alumnos. Se totalizan las respuestas correctas e incorrectas.

	ÍTEMS			
	1	2		20
Alumno 1	1	0		0
Alumno 2	0	0		1
Alumno 3	1	1		1
Alumno 4	1	1		1
Alumno 40	1	0		1
Correctas	10	24		30
Incorrectas	30	16		10

Grado de dificultad del ítem Nº 1, respondido correctamente por 10 de los 40 alumnos:

$$Id_1 = \frac{10}{40} = 0{,}25\% \text{ ó } 25\% \text{ de dificultad}$$

Grado de dificultad del ítem Nº 20, respondido correctamente por 30 de los 40 alumnos:

$$Id = \frac{30}{40} x\, 100 = 75\%$$

Estas cifras se interpretan según la siguiente tabla:

85% de respuestas correctas	=	Muy fáciles
50 al 85% de respuestas correctas	=	Relativamente fáciles
15 al 50% de respuestas correctas	=	Relativamente difíciles
0 a 15 % de respuestas correctas	=	Difíciles o muy difíciles.

25 % significa que el ítem es relativamente difícil.

76 % significa que el ítem es relativamente fácil.

16. ÍNDICE DE DISCRIMINACIÓN DE UN ÍTEM

Un buen ítem debe discriminar entre los alumnos que realmente conocen la repuesta y los que no la conocen. El índice de discriminación de un ítem indica el número de discriminaciones que el ítem puede producir. Se basa en la relación entre una respuesta y el puntaje total de la prueba, presuponiéndose que ésta realmente mide lo que pretende medir y que en consecuencia el mejor ítem será aquel cuya correlación con el puntaje total sea más alta.

El índice de discriminación de un ítem resulta de hallar la diferencia entre la proporción del grupo superior que respondió bien el ítem y la proporción del grupo inferior que también respondió bien.

Para determinar el grupo superior e inferior se consideran las puntuaciones totales obtenidas por los estudiantes a quienes se les aplicó el test. Aquellos que obtuvieron los más altos porcentajes son considerados como los más conocedores y, por tanto, constituyen el grupo superior y los que obtuvieron los más bajos, como los menos conocedores y constituyen el grupo inferior. Ejemplo: De 100 alumnos que respondieron una prueba, los 27 alumnos que obtuvieron puntuaciones más altas serán considerados miembros del grupo superior y los 27 que obtuvieron puntuaciones más bajas, serán considerados miembros del grupo inferior.

Averiguaremos luego cuántos alumnos de los 27 del grupo superior respondieron el ítem correctamente y cuántos alumnos de los 27 del grupo inferior también lo hizo correctamente. Luego se obtiene la diferencia.

Si los 27 alumnos del grupo superior respondieron correctamente el ítem, o sea el 100%, y en el grupo inferior sólo 12 de los 27 también lo hicieron correctamente, o sea el 44%, estableciendo la diferencia tenemos: 100 % - 44 % = 56%, que es el índice de discriminación del ítem analizado.

El índice de discriminación puede hallarse también aplicando la siguiente fórmula:

$$Id = \frac{pms - pmi}{N} x\,100$$

Donde:

Id = Índice de discriminación.

pms = puntaje máximo alcanzado por el grupo superior.

pmi = puntaje máximo alcanzado por el grupo inferior.

N = número de alumnos.

Ejemplo: Hallar el índice de discriminación de un ítem cuyas respuestas correctas del grupo superior son 25 y las respuestas correctas del grupo inferior son 12. El número de examinados en los dos grupos es 50.

Reemplazando los valores se tiene:

$$Id = \frac{pms - pmi}{N} x\,100$$

La interpretación de este resultado se hace empleando la siguiente tabla:

40% a más = Alto índice de discriminación.

30 al 39% = Índice de discriminación bueno.

20 al 29% = Índice de discriminación regular.

menos del 19% = Deficiente índice de discriminación.

La cifra hallada, 26 %, indica que el índice de discriminación de este ítem es regular.

Otro ejemplo: Se ordena en forma descendente el promedio de las notas.

	ÍTEMS				
	1	2	3	....	20
Alumno 1	1				0
Alumno 2	0				1
...	...				...
Alumno 20	1				1
Sub Total	16				**17**
Alumno 21	0				0
Alumno 22	1				1
...					
Alumno 40	0				0
Sub Total	9				**3**

Se totaliza el número de respuestas correctas de cada ítem y por cada grupo. En este caso, el grupo superior es 20 y el inferior es 20. El número total de alumnos es 40.

Se ha ordenado a los estudiantes en función de su rendimiento académico, de mayor a menor. Para hallar el índice de discriminación del ítem Nº 1 se hacen los siguientes cálculos:

$$I_d = \frac{16-19}{40} = \frac{7}{40} = 0,175 = 17,5\%$$

Según la tabla, este es un ítem deficiente y debe descartarse.

Para obtener el índice de discriminación del ítem 20 se procede del siguiente modo:

$$Id = \frac{15-3}{40} = \frac{12}{40} = 0,3 = 30\%$$

Este ítem, con un índice de discriminación de 30%, es bueno, aunque podría mejorarse.

LECTURAS COMPLEMENTARIAS

FUNDAMENTOS DE MEDICIÓN

Donald Ary*

Uno de los propósitos de la investigación educacional es lograr un entendimiento más profundo de las relaciones que hay entre las variables de las poblaciones. Por ejemplo, se podría preguntar: ¿qué relación hay entre la inteligencia y la creatividad de los niños de seis años? No es posible observarlas directamente, ni a todos los niños de esa edad. Pero esto no quiere decir que debamos permanecer en la ignorancia acerca de éstas y otras preguntas similares. Se dispone de índices que ofrecen aproximaciones de las construcciones, inteligencia y creatividad; es decir, hay comportamientos observables que son aceptados como indicadores válidos de ambas. El uso de índices para lograr aproximaciones de las construcciones constituye el aspecto cuantitativo de la investigación.

INSTRUMENTOS DE MEDICIÓN

Es importante seleccionar instrumentos seguros de medición con el propósito de cuantificar los comportamientos y atributos que habrán de estudiarse. En algunos casos esto no presenta ningún problema. Por ejemplo si se desea indagar el orden de nacimiento de un grupo de sujetos o la escolaridad de sus padres, lo único que hay que hacer es plantearles estas preguntas a los sujetos y registrar sus respuestas. Sin embargo, la tarea de cuantificar la información no es siempre tan sencilla.

Algunas interrogantes requieren la creación de instrumentos confiables que midan cualidades abstractas y complicadas. Habrá que seleccionar o idear escalas e instrumentos capaces de medir características tales como inteligencia, aprovechamiento, personalidad, motivación, actitudes, aptitudes, intereses y otras. La cuantificación de cada cualidad exige un instrumento adecuado. En la siguiente sección expondremos brevemente algunas de las clases de instrumentos que se utilizan en la investigación educacional.

---

* Tomado de: ARY, Donald et al.: *Introducción a la investigación pedagógica*. México, D.F. McGraw-Hill Interamericana, 1989. pp. 178-202.

Entrevistas y cuestionarios

Una forma de obtener datos consiste simplemente en hacer preguntas. Las entrevistas y los cuestionarios aplican este método. Con ellos se recoge información sobre hechos, creencias, sentimientos, intenciones, etc. Aunque en ambos se emplean preguntas, muestran notables diferencias.

Entrevistas

Las entrevistas son un método muy útil para recabar datos y gracias a ciertas cualidades especiales todavía se usan muchísimo. La flexibilidad es uno de sus aspectos más sobresalientes. El "rapport" (relación personal) que se establece con los sujetos crea una atmósfera de cooperación en la que puede obtenerse información verídica. Hay que tomar en cuenta la clase de persona que se va a entrevistar y la situación dentro de la cual se desarrolla la entrevista. El entrevistador puede explayarse sobre alguna pregunta, explicar su significado en caso de que a su interlocutor le parezca poco clara. Ninguno de los otros tipos de recopilación de datos, entre ellos los cuestionarios y las pruebas, ofrece estas ventajas.

Hay dos tipos de entrevistas: *estructuradas y no estructuradas.* En las primeras las preguntas y las respuestas alternativas que están permitidas a los sujetos se fijan de antemano y se aplican con rigidez a todos ellos. La ventaja es que se trata de un sistema normalizado, de modo que las respuestas pueden ser clasificadas y analizadas fácilmente. Su desventaja es la falta de flexibilidad y que dan la impresión de ser demasiado formales. Las restricciones impuestas a este tipo de entrevista aumentan su confiabilidad, pero pueden mermar su profundidad.

Las entrevistas no estructuradas son más informales. Es posible interrogar libremente a los sujetos sobre sus opiniones, actitudes, creencias y otros puntos. Estas entrevistas son flexibles y suelen planearse para adaptarse a los sujetos y a las condiciones dentro de las que se desarrollan. Los sujetos tienen libertad para ir más allá de las simples respuestas y expresar sus puntos de vista en la forma que deseen. Las preguntas pueden desviarse de los planes originales y centrarse en los puntos que parezcan importantes. Todo eso requiere personal experto, capacitado y alerta.

No se prestan mucho a cuantificación, pero ayudan a generar y aclarar las dimensiones presentes en la materia que se estudia. Se utilizan mucho en la psicología clínica, en el asesoramiento, en la dirección y en el estudio de casos. En la investigación su empleo suele limitarse a las etapas preliminares, en las que el investigador escoge las variables del estudio. Por ejemplo, para determinar las habilidades que los patrones potenciales desean que se enseñen en un programa de capacitación para el trabajo de oficina, podría comenzarse con entrevistas no

estructuradas destinadas a identificar cuáles habilidades se mencionan con mayor frecuencia, enumerarlas y solicitar a los patrones, mediante entrevistas estructuradas, que estimen la importancia de cada una.

Cuestionarios

En las entrevistas el contacto directo con los sujetos requiere tiempo y dinero. Una buena parte de esa información puede obtenerse por medio un cuestionario escrito. Si se le compara con la entrevista, ésta es más eficiente y práctica y permite usar una muestra más amplia. En la investigación experimental se utiliza frecuentemente.

He aquí otra de sus ventajas: los sujetos reciben instrucciones normalizadas y ni la apariencia personal ni el estado de ánimo ni la conducta del examinador influyen en los resultados finales.

Los cuestionarios son de dos clases: *estructurados,* o de forma cerrada, y *no estructurados,* o de forma abierta. Los primeros contienen preguntas y respuestas alternativas. Las respuestas dadas a cada pregunta deberán agotar todas las posibles preguntas y excluirse mutuamente. En los cuestionarios no estructurados no hay respuestas sugeridas.

La administración y la calificación de un cuestionario estructurado se realiza en forma directa, y los resultados se prestan a un análisis fácil.

Este tipo de cuestionario tiene la desventaja de obligar a los respondientes a escoger entre una de varias respuestas alternativas seleccionadas previamente, aunque tal vez no vislumbren respuestas claras, o a elegir alternativas que no representen realmente sus actitudes.

Por el contrario, los cuestionarios no estructurados tienen la ventaja de dejarlos en libertad para que expresen sus opiniones y sus actitudes, sólo que la información que aportan es difícil de procesar y analizar. Al contestar cuestionarios no estructurados, los sujetos pueden omitir puntos importantes, recalcar asuntos que carezcan de interés para el investigador o que no tengan importancia para el estudio. Por tal razón la mayoría de los investigadores prefiere los cuestionarios estructurados.

Una desventaja de ambos tipos es la posibilidad de que los respondientes malinterpreten las preguntas. Es muy difícil elaborar una serie preguntas cuyo significado sea tan claro como el agua para todos. El investigador puede saber con exactitud lo que significa una pregunta, pero por una formulación deficiente o la distinta acepción que se da a los términos, el sujeto los entiende erróneamente. Esta desventaja puede obviarse mediante un estudio piloto en el que se ensaya el cuestionario con unos cuantos sujetos, representativos de aquellos a quienes se aplicará.

Una desventaja peculiar del envío de cuestionarios por correo en su bajo porcentaje de respuestas. Por lo común cabe esperar que regresen menos de la mitad, lo cual no sólo reduce el tamaño de la muestra sino que puede viciar los resultados, e impedir que se hagan generalizaciones válidas. No es legítimo suponer que la ausencia de respuestas se distribuye aleatoriamente a lo largo de un grupo.

Los estudios han mostrado que suele haber diferencias sistemáticas en las características de los sujetos que responden o no a los cuestionarios. Los que contestan pueden ser más inteligentes, educados, concienzudos e interesados en la cuestión y en general más favorables al tema del cuestionario. Una muestra no representativa como ésta puede dañar la generalización de los resultados. La meta de un cuestionario es obtener de un 70 a 80 por ciento de respuestas. Si la ausencia de respuestas supera al 30 por ciento, el valor de los resultados se pondrá en duda.

Un cuestionario bien construido tiene mayores probabilidades de producir una buena respuesta que uno que esté mal elaborado.

Elaboración de cuestionarios

Es una tarea difícil y que requiere tiempo. A continuación presentamos algunas indicaciones para redactar los reactivos de un cuestionario escrito.

1. *Construir el instrumento de modo que refleje calidad.* Un cuestionario que parezca haber sido integrado sin orden y con mucha rapidez no obtendrá un alto porcentaje de respuestas. Durante el proceso de elaboración a veces se necesitan muchas revisiones para eliminar reactivos ambiguos o innecesarios.
2. *Hacer el cuestionario lo más breve posible, de modo que se requiera el mínimo de tiempo para llenarlo.* Es más probable que los sujetos contesten y regresen un cuestionario corto. El examinador debe suprimir todos los puntos innecesarios, en especial aquellos cuyas respuestas puedan hallarse en otras fuentes. Todos los reactivos deberán estar en función del problema de investigación; es decir, habrán de aportar datos necesarios para comprobar la hipótesis o contestar las interrogantes. Por ejemplo, puede eliminarse una pregunta sobre la edad del respondiente en un estudio donde esta información no sea necesaria para el análisis de la mayoría de los datos.
3. *Redactar los reactivos del cuestionario de manera que todos puedan entenderlo.* Se utilizará un lenguaje que no sea técnico y que se ajuste al respondiente menos educado. Las oraciones serán cortas y simples. Conviene hacer que otras personas, que de preferencia tengan una educación similar a la de las que habrán de participar en el estudio, lean e interpreten el contenido de cada pregunta. Por ejemplo, las interrogantes que incluyan términos como, *gasto*

deficitario, balanza de pagos, producto nacional bruto, no serán apropiadas en una encuesta dirigida al público en general.

4. *Formular el cuestionario de modo que produzca respuestas que no sean ambiguas.* Cuando sea posible, las respuestas deberán estar cuantificadas. Por ejemplo, en lugar de que los respondientes marquen "*algunas veces*", "*con frecuencia*" o "*siempre*", exprésese la alternativa "*número de veces por semana*". Es decir, la redacción de las preguntas no deberá tener distintos significados para la gente.

5. *Elaborar las preguntas evitando los prejuicios que pudieran influir* u orientar al sujeto hacia determinada dirección. Por tal razón conviene evitar las palabras estereotipadas que impliquen prestigio, las cargadas de emoción o superlativas.

   Por ejemplo, la pregunta ¿se ha inscrito usted en el registro federal de electores para votar? Sería preferible a decir: ¿ha ejercido usted sus derechos ciudadanos registrándose para votar? La pregunta ¿estaría usted de acuerdo en que Estados Unidos desarrollase el potencial atómico como una fuente alternativa de energía? Podría producir diferentes respuestas a las que originaría que esta otra: En razón de la huelga del carbón más larga de la historia, ¿recomendaría usted que Estados Unidos desarrollase el potencial atómico como una fuente alternativa de energía?

6. *Los reactivos de un cuestionario no deberán ser engañosos a causa de supuestos implícitos.* El marco de referencia para responder las preguntas deberá ser claro y uniforme para todos los sujetos. Si hay que plantear alguna premisa antes de que den la respuesta, deberán incluirse preguntas tendientes a indagar esas suposiciones.

   Por ejemplo, en una encuesta diseñada para estudiantes del último año, la pregunta siguiente: ¿Piensa usted que la enseñanza superior lo ha preparado adecuadamente para la universidad?, supone que el estudiante asistirá a la facultad y que está al tanto de la preparación requerida. Esta otra: ¿Se ha registrado usted para votar en la próxima elección?, supone que sea mayor de 18 años, lo cual tal vez no sea así.

7. *Las alternativas de las preguntas del cuestionario deben ser exhaustivas; es decir, se expresarán todas las alternativas posibles sobre la cuestión.* Por ejemplo, ¿cuál es su estado civil? Es una pregunta que debe incluir no sólo las alternativas *casado* o *soltero,* sino también *viudo, divorciado, separado.*

   Al idear las alternativas para cuestionarios cuyo fin es identificar actitudes u opiniones sobre ciertos temas, es conveniente presentar primero las preguntas

en forma abierta a una muestra pequeña de sujetos. Sus respuestas pueden utilizarse después como alternativas en la redacción final. En las preguntas con gran variedad de posibles respuestas, inclúyase siempre la alternativa *otra* acompañada de una petición para que el sujeto explique su elección. Por ejemplo, la pregunta: ¿Cuál es su posición en el sistema escolar? deberá comprender las alternativas *administrador, maestro, bibliotecario, otra.*

8. *Evitar preguntas que pudieran ocasionar reacciones de turbación, desconfianza u hostilidad.* Las preguntas no deberán de poner a la defensiva. Por ejemplo, a la gente no le gusta que le hagan preguntas acerca de su edad, ingresos, religión o nivel educacional. Es preferible escribir una pregunta relacionada con la edad en términos de escalas —en vez de edades exactas— pues con ello se les permite elegir la categoría apropiada. La pregunta: *¿Tiene usted un diploma de enseñanza media?,* podría causar vergüenza a quien no la haya terminado. La pregunta debería ser: *¿A qué grado había usted llegado cuando abandonó la escuela?*

9. *Colocar las preguntas en correcto orden psicológico. Cuando se planteen preguntas generales y específicas sobre un mismo tema, presentar las generales primero.* Por ejemplo, verifíquese si los sujetos están satisfechos o no con sus condiciones laborales antes de preguntarles sobre los cambios que recomendarían. Las preguntas objetivas sobre un tema o situación han de preceder a las más subjetivas. Algunas veces los respondientes se muestran reticentes a contestar preguntas acerca de actitudes, preferencias, sentimientos personales, motivos y cosas por el estilo. Pero si con preguntas objetivas se aclara y especifica la situación, les será tal vez más fácil formular y expresar sus opiniones sobre la cuestión.

10. *Disponer las preguntas de manera que las respuestas puedan tabularse e interpretarse inmediatamente.* Es necesario prever la forma de los datos necesaria para el análisis y asegurarse que el cuestionario producirá esos datos.

11. *El cuestionario debe ir acompañado de una carta explicatoria, con el nombre y la profesión del respondiente.* La carta explica el propósito y el valor del estudio y la razón por la que el sujeto fue incluido en la muestra. Por lo tanto deberá motivarlo para que conteste. Es preciso hacerle creer que va a hacer una contribución importante al estudio. La firma de la carta también puede influir en la devolución del cuestionario. Será de gran utilidad que una persona muy conocida por los sujetos, como el director de un departamento universitario o el decano de una escuela, firme esa carta. Esta tendrá probablemente mayor efecto que la de un estudiante graduado desconocido. Si existe un patrocinador, digamos alguna institución o agencia, se le mencionará.

Conviene dejarles la posibilidad de permanecer en el anonimato. En cualquier caso, debe asegurárseles que sus respuestas serán confidenciales. También hay que ofrecerles compartir los resultados del estudio con ellos si están interesados. Es necesario adjuntar un sobre con estampillas y un sobre con el domicilio del examinador. Si los sujetos permanecen en el anonimato es aconsejable incluir una tarjeta postal, que será mandada por separado al investigador y que indique que el cuestionario ha sido depositado ya en el correo. De esta manera puede llevarse un registro de los cuestionarios devueltos.

Pedir una contestación inmediata del cuestionario. Si se concede un lapso de tiempo, por ejemplo, de un mes, los sujetos pueden darle largas al cuestionario y olvidarlo a pesar de sus buenas intenciones.

12. *Un recordatorio es necesario si se desea alcanzar el porcentaje máximo de contestaciones*. Cuando el cuestionario no haya sido contestado inmediatamente después de haberlo mandado, envíese a los sujetos un recordatorio en forma de tarjeta postal. Hecho eso se recomienda mandar por segunda vez el cuestionario acompañado de otra carta explicatoria. Tal vez deba usarse un tono distinto para persuadir a los respondientes a llenar y regresar el cuestionario. Si el lector quiere obtener mayor información sobre la elaboración de cuestionarios, consulte un texto de Selltiz, Wrightsman y Cook.

Cómo resolver el problema de ausencia de respuestas

El método usual consiste en tratar de entrevistar a una pequeña muestra aleatoria de los que no contestaron, con el propósito de saber algo sobre sus características y lograr sus respuestas. Quizás se averigüe que no hay ninguna diferencia entre sus respuestas y las de los sujetos originales. En este caso es posible pensar que los que contestan representan una muestra imparcial de todos los que recibieron el cuestionario. Después de establecer el patrón de respuesta mediante las entrevistas con ellos, es posible ponderar sus respuestas en el análisis final de los datos.

Si se descubre que ciertos subgrupos claramente identificables no contestaron el cuestionario, posiblemente haya que cambiar la interrogante de investigación para excluirlos. Por ejemplo, si los maestros de secundaria mostraron un porcentaje mucho menor de devoluciones que los de primaria, el investigador podría restringir la encuesta en estos últimos y formular de nuevo la pregunta para señalarlo.

Validez de los cuestionarios

En la investigación, con mucha frecuencia se utilizan cuestionarios sin reflexionar sobre cuán adecuados son para hacer las mediciones que se desean. Esto se relaciona con el problema de la validez. Es muy importante señalar que la validez no puede suponerse sino que debe probarse.

Un factor que afecta de manera especial a la validez de los cuestionarios es el hecho de que se requiera una firma o no. Hay buenas razones para creer que se obtendría mayor veracidad si los respondientes pudieran permanecer en el anonimato. No obstante, esta probabilidad depende de la naturaleza de las preguntas incluidas es decir, si se busca o no una información muy personal. Si la identificación de ello no es importante para el análisis de resultados convendrá ofrecerles el anonimato.

Algunos estudios han usado la observación directa del comportamiento como criterio de validez de las respuestas en un cuestionario. Es decir, después de conseguir muchas respuestas mediante el cuestionario, se realizan observaciones para confirmar si el comportamiento real de los sujetos concuerda con las actitudes y opiniones que expresaron. En otros casos las consideraciones sobre la validez requieren estudios sobre la interpretación que hagan de las preguntas. Tales estudios suelen revelar ambigüedades o imprecisiones de las preguntas.

Pruebas

Las pruebas (tests) constituyen valiosos instrumentos de medición en el análisis educacional. Son una serie de estímulos que se presentan a un individuo para suscitar respuestas, en base a las cuales se asigna una puntuación numérica. Esta puntuación se apoya en una muestra representativa del comportamiento individual, y es un índice del grado en que el sujeto posee las características que se van a medir. Dos requisitos esenciales de los tests son la validez y la confiabilidad.

Otra condición indispensable de las pruebas es su objetividad, que se muestra por un nivel máximo de concordancia entre los calificadores. Una vez preparada la clave de registro para una prueba objetiva, cualquier persona aunque no esté entrenada, o incluso una máquina, pueden calificar el test.

Con las pruebas objetivas se puede examinar gran variedad de características. Algunas fuentes de utilidad para encontrar pruebas que se amolden a los propósitos de una investigación son *Mental Measurements Yearbooks* de Buros y *Tests in Print II.* En los manuales los editores de pruebas incluyen las instrucciones para administrarlas, calificarlas e interpretarlas

Pruebas de aprovechamiento

En la mayoría de las investigaciones sobre la eficacia de los métodos didácticos la variable dependiente es el aprovechamiento. De ahí que las pruebas de aprovechamiento se utilicen ampliamente en la investigación educacional así como en los sistemas escolares. Miden el dominio y la destreza en distintas áreas de conocimiento.

Se clasifican en estandarizadas y aquellas que son construidas por el maestro o el investigador. Las primeras son pruebas publicadas y son fruto de una preparación cuidadosa y experta, que cubren amplios objetivos académicos comunes a gran número de sistemas escolares. Estas pruebas disponen de normas comparativas; su validez y confiabilidad se ha probado; se cuenta con instrucciones para administrarlas y calificarlas.

Con el propósito de establecer las normas que las rigen, sus creadores las aplicaron a una muestra que escogieron para representar al país en su totalidad.

La media de un nivel de grados particular se convierte en la norma para ese nivel. Las habilidades medidas no constituyen necesariamente lo que "sería necesario" enseñar en cualquier grado, pero el uso de normas suministra una base a los investigadores para comparar sus grupos con una estimación de la media para todos aquellos alumnos que se hallan en ese nivel.

Hay pruebas estandarizadas para asignaturas individuales, como matemáticas o química, y también en forma de baterías generales que miden diversas áreas de aprovechamiento. Por ejemplo, la Prueba California de Aprovechamiento (CAT) contiene exámenes en las áreas de lectura, lenguaje y aritmética. Las pruebas consecutivas del progreso educacional (STEP) cubren siete áreas.

Al seleccionar una prueba de rendimiento los investigadores tienen que ser muy cuidadosos y elegir una que sea confiable y resulte apropiada (válida) para medir el aspecto del rendimiento en el cual estén interesados. Algunas veces les será imposible escogerla pues tendrán que usar la que el sistema escolar haya seleccionado. Buros presenta una lista muy amplia y reseñas de distintas pruebas de aprovechamiento disponibles.

Si una prueba existente mide el comportamiento deseado y si la confiabilidad y las normas resultan adecuadas para el propósito del estudio, hay ventajas en el uso de un instrumento estandarizado. Aparte del tiempo y esfuerzo que se ahorran, la continuidad de los procedimientos de comprobación constituye ya una ventaja. Es decir, usando el mismo instrumento se pueden comparar los resultados de los estudios e interpretarse con respecto a los de otros estudios.

Si las pruebas estandarizadas de aprovechamiento no se consideran adecuadas para los objetivos de un estudio de investigación, los experimentadores pueden construir sus propias pruebas. Es preferible hacer esto que emplear alguna prueba estandarizada que resulte inadecuada aunque se encuentre disponible.

En este caso deberá prepararse la prueba con gran cuidado, particularmente al determinar su validez y confiabilidad antes de aplicarla. Algunas indicaciones sobre la construcción de pruebas se hallan en textos especializados en medición como los de Gronlund, Stanley, Thorndike y Hagen.

Algunas veces los promedios de las calificaciones de los sujetos en sus escuelas se utilizan como indicadores de éxito y aprovechamiento académico. Cuando participen estudiantes de distintos sistemas escolares deberán tomarse ciertas precauciones. Las letras de calificación de diferentes escuelas no significan necesariamente lo mismo, y por tanto no indican el mismo grado de dominio y destreza.

Pruebas de inteligencia

Mientras que las pruebas de aprovechamiento intentan medir el rendimiento en áreas específicas, las de inteligencia se ocupan del rendimiento general. Se proponen medir la capacidad del sujeto para percibir relaciones, resolver problemas y aplicar el saber en una variedad de contextos.

No se les ha de considerar como mediciones de la inteligencia " pura" o innata. El rendimiento en esta clase de pruebas depende en parte de la experiencia y de los antecedentes del sujeto. La controversia sobre el significado del concepto inteligencia y la tendencia de algunos a asociarla con habilidad heredada, ha hecho que el uso del término *inteligencia* para describir estas pruebas haya disminuido en los últimos años. Paulatinamente se la ha ido substituyendo por *aptitud académica* –locución más descriptiva ya que señala la función principal de esas pruebas: predecir el aprovechamiento académico.

Estas pruebas son de gran utilidad para los educadores y poseen una validez muy amplia para predecir el éxito escolar. Los investigadores también las usan mucho. La inteligencia es una variable independiente que debe ser controlada en los experimentos educacionales. Y para ello el investigador utiliza las puntuaciones de alguna prueba de inteligencia. Algunas de ellas han sido diseñadas para aplicarlas a individuos y otras para usarlas con grupos.

Pruebas individuales de inteligencia.

Las más conocidas son las de Stanford-Binet y las tres pruebas de Wechsler. La Stanford-Binet actualmente en uso es el resultado de varias revisiones hechas a la que creó Alfred Binet en Francia para medir las diferencias en la madurez mental de los niños. En su forma original servía para medir la *edad mental.*

Más tarde apareció el concepto de *cociente intelectual (CI),* que se obtiene dividiendo la edad mental (EM) entre la edad cronológica (EC) y multiplicando el resultado por 100. La presente versión produce edades mentales pero no emplea la razón EM/EC para determinar el CI. El cociente intelectual se obtiene comparando el rendimiento de un individuo (la puntuación) con las normas obtenidas de su grupo

cronológico mediante el uso de puntuaciones estándar. Una característica importante de este tests consiste en que da una medición general de la inteligencia. No intenta medir habilidades separadas, como lo hacen otras pruebas.

Las pruebas Wechsler son el producto de un intento de David Wechsler por crear un instrumento para medir la inteligencia de los adultos. Su primera prueba se publicó en 1939 y le siguieron otras tres: una en 1949, la Escala Wechsler de Inteligencia para Niños (WISC), que fue revisada en 1974 y se conoce como la WISC-R; otra en 1955, la Escala Wechsler de Inteligencia para Adultos (WAIS); y la última en 1967, la Escala Wechsler de inteligencia para Niños de Nivel Preescolar y de Primaria (WPPSI), que se implantó para el grupo de edad comprendido entre 4 y 6 años y medio.

Un aspecto destacado de las pruebas Wechsler es que están divididas en subpruebas, lo cual le permite al investigador obtener dos puntuaciones para cada sujeto, una para el CI verbal y otra del CI no verbal. Estas subpruebas se dividen además en subescalas, que indican el rendimiento del sujeto en tareas específicas.

Pruebas colectivas de inteligencia.

Las pruebas de Stanford-Binet o de Wechsler las debe aplicar un psicometrista experto a un sujeto individual, lo cual comporta un gasto de tiempo y dinero. Son imprácticas cuando se desea conseguir mediciones de inteligencia para grandes grupos. En estos casos se emplean pruebas colectivas. La primera prueba colectiva de la capacidad mental fue creada durante la Primera Guerra Mundial para medir la inteligencia de los hombres que prestaban servicio militar. Una versión de esta prueba, la Army Alpha, comenzó a aplicarse en la vida civil después de la guerra y se convirtió en el modelo de muchas pruebas colectivas. Hoy día existe un número considerable de este tipo de tests.

Inventarios de Personalidad

Obtener mediciones de la personalidad constituye otra área de interés para los pedagogos. Hay diversos tipos de mediciones de la personalidad, cada uno de los cuales refleja un distinto punto de vista teórico. Algunos provienen de las teorías de los rasgos y de la tipología, mientras que otras emanan de las teorías psicoanalíticas y de la motivación. Los investigadores deben saber con precisión lo que desean medir y seleccionar después el instrumento, prestando gran atención a los datos relativos a su validez. Los tres tipos de medición de la personalidad más comunes en la investigación son los siguientes: 1) inventarios, 2) escalas de clasificación, 3) técnicas proyectivas.

Inventarios

En un inventario se presenta a los sujetos una recopilación extensa de enunciados que describen normas de comportamiento; se les pide decir si cada enunciado es característica o no de su conducta marcando *si*, *no* o *indeciso.* Las puntuaciones se calculan contando el número de respuestas que concuerden con el rasgo que el investigador está tratando de medir. Por ejemplo, cabe esperar que los paranoicos contesten afirmativamente a este enunciado: *Las personas siempre andan murmurando a mis espaldas*, y que respondan negativamente a este otro: *Yo pienso que la policía debe ser justa y razonable*. Por supuesto si tales respuestas se dan solamente a dos incisos nos indicarán tendencias paranoicas. Sin embargo, podrá verse en ellas un indicador de paranoia si se encuentran en varios reactivos.

Algunos de los inventarios introspectivos miden un solo rasgo, como la *EscalaF de California*, que se ocupa del autoritarismo. Otros, como el cues*tionario de los dieciséis factores de la personalidad* de Cattell, miden varios rasgos. He aquí algunos de los más usados en la investigación: el *Inventario multifásico de la personalidad de Minnesota,* la *Encuesta de la preferencia de temperamento de Guilford-Zimmerman,* la *Lista de verificación de problemas de Mooney* y el *Programa de preferencias personales de Edwards.*

Los inventarios pueden usarse en la investigación educacional para obtener descripciones de las características de ciertos grupos definidos, como los estudiantes de bajo rendimiento, los que abandonan sus estudios, los miembros de grupos minoritarios, etc. También se utilizan en estudios sobre las interacciones de los rasgos de la personalidad y de variables como inteligencia, aprovechamiento y actitudes.

Los inventarios poseen las ventajas de economía, sencillez y objetividad. La mayoría de sus desventajas están relacionadas con el problema de la validez.

Ésta depende en parte de la capacidad de sujetos para leer y entender los reactivos, su conocimiento de sí mismos y especialmente su buena disposición para dar respuestas francas y honestas. De ahí que la información de los inventarios pueda ser superficial o viciada. Dicha posibilidad debe tenerse en cuenta al emplear los resultados obtenidos con ellos.

Escalas de clasificación

Uno de los instrumentos de medición más usado es la escala de clasificación. Consiste en que una persona evalúe el comportamiento o rendimiento de otra. Generalmente se le pide al clasificador colocar al sujeto en algún punto de un continuo o en una categoría que describa su comportamiento típico. Un valor numérico se

asigna al punto o categoría. Se supone que los clasificadores están familiarizados con el comportamiento típico del individuo.

La clasificación se ha venido utilizando en investigaciones sobre el desarrollo de los niños y en muchos otros aspectos de la conducta.

Hay diversos tipos de escalas de clasificación. Una de las más comunes es la *Escala gráfica*, en que el clasificador se limita a colocar una marca en el punto apropiado de una línea horizontal que une ambos extremos de la conducta. La figura 1 contiene un ejemplo de esta clase de escalas. El clasificador puede marcar cualquier punto de la línea continua. En algunas escalas gráficas el constructor de la prueba asigna valores numéricos a los puntos descriptivos, lo cual origina una *Escala de clasificación numérica*.

EJEMPLO DE UNA ESCALA GRÁFICA

	baja	*media*	*alta*
Apariencia Personal			
Aceptabilidad social			
Habilidades oratorias			

Figura 1

El reactivo sobre las habilidades oratorias de la figura podría aparecer de la siguiente manera en una escala numérica:

1	2	3	4	5	6	7
Uno de los oradores más mediocres			Un orador medio			Uno de los mejores oradores

Un segundo tipo de escala de clasificación es la de categorías, que comprende un número de categorías dispuestas en una serie ordenada. La serie más usada abarca de 5 a 7 categorías. El clasificador escoge la que representa mejor el comportamiento de la persona que va ser estudiada. Supongamos que se estiman las habilidades de un estudiantes y que una de las características por clasificar es la creatividad. Un reactivo podría ser éste:

¿Cuán creativa es esta persona? (marcar uno)

Excepcionalmente creativa.
Muy creativa.

Creativa.

No creativa.

Absolutamente no creativa.

Algunas veces las categorías en este tipo de escalas constan de breves frases descriptivas. Por ejemplo:

¿Cuán creativa es esta persona?

Siempre tiene ideas creativas.

Tiene muchas ideas creativas.

Algunas veces produce ideas creativas.

Al usar las escalas gráficas y de categorías, los estimadores emiten sus juicios sin comparar directamente al sujeto con otros individuos o grupos. En las *Escalas de clasificación comparativas,* por otra parte, se les instruye para que hagan sus juicios con referencia directa a las posiciones de otros sujetos con las cuales podría compararse el individuo. Las posiciones de la escala de clasificación se definen de acuerdo con una población con características conocidas. En la figura 2 se muestra una de las escalas. Esta podría utilizarse para seleccionar aspirantes a una escuela universitaria de graduados. Al estimador se le solicitaría que juzgase la habilidad del aspirante para realizar actividades propias de graduados en comparación con todos los estudiantes que conoce. Para que su estimación tenga validez, debe estar al tanto de la gama y distribución de las habilidades del grupo total de estudiantes graduados.

Todas las técnicas de clasificación están sujetas a errores considerables, lo cual reduce su validez y confiabilidad. Entre los errores sistemáticos más frecuentes en que se incurre al clasificar personas figura el *efecto del halo,* que aparece cuando los estimadores permiten que una impresión generalizada del sujeto influya sobre la clasificación de algunos aspectos específicos del comportamiento. Esta impresión se extiende de un reactivo de la escala a otro. Por ejemplo, un maestro podría clasificar a un estudiante que realiza una buena actividad académica como superior en inteligencia, popularidad, honestidad, perseverancia y en todos los otros aspectos de la personalidad.

Existe también el *error de generosidad,* o sea la tendencia a conceder al sujeto el beneficio de la duda. Es decir, cuando los estimadores se muestran inseguros es probable que se expresen favorablemente de la persona. En cambio el *error de severidad* es la inclinación a dar a los individuos una clasificación muy baja en todas las características. El *error de la tendencia central* es la propensión a evitar cualquier extremo y situarlos en la mitad de la escala.

EJEMPLO DE UNA ESCALA DE CATEGORÍAS

Área de competencia (que habrá de clasificarse)	Excesivamente bajo	Peor que la mayoría de los estudiantes	Cerca del promedio estudiantil	Mejor que la mayoría	Muy superior	Imposible de juzgar
1. ¿Muestra esta persona tener metas profesionales bien definidas y notables?						
2. ¿Afronta los problemas en forma constructiva?						
3. ¿Toma en cuenta las críticas positivas y las usa en forma constructiva?						

Figura 2

Una forma de evitar estos errores consiste en entrenar a fondo a los estimadores antes que entren en acción. Deberán estar informados sobre la posibilidad de cometer uno de estos errores. Y es de gran importancia que dispongan de tiempo suficiente para observar al sujeto y a su comportamiento antes de llevar a cabo la clasificación.

Otra forma de prevenir un error es asegurarse que el comportamiento que se va a estimar y los reactivos de la escala estén definidos con claridad. Los reactivos deberán describirse en términos del comportamiento total observable y no de las conductas que obliguen al estimador a realizar inferencias. Una exposición notable sobre la forma de evitar el error del estimador se halla en la obra de Guilford.

La confiabilidad de este procedimiento suele incrementarse al hacer que varios estimadores se ocupen de un mismo sujeto de forma independiente. Con ello se consiguen clasificaciones independientes que se combinan o prorratean para obtener una puntuación total.

Técnicas proyectivas

Son medidas en las cuales se pide a un individuo que responda a estímulos ambiguos o no estructurados. Reciben ese nombre porque se espera que el sujeto

proyecte dentro del estímulo sus necesidades, deseos, temores, ansiedades, etc. Basándose en las interpretaciones y respuestas, el examinador intenta construir un cuadro general de la estructura de la personalidad del individuo.

Los métodos proyectivos son usados principalmente por los psicólogos clínicos para el estudio y el diagnóstico de los que sufren problemas emocionales. En la investigación educacional, se usan poco porque se necesita un entrenamiento especial para administrarlas y calificarlas y del gasto que representa su aplicación a un individuo aislado. Además, algunos investigadores creen que su validez no ha sido probada satisfactoriamente.

Las dos técnicas proyectivas más conocidas son la prueba de Rorschach y la de Apercepción Temática (TAT). La primera utiliza manchas de tinta como el estímulo; en la segunda se le muestran al sujeto varias fotografías y se le pide que invente una historia acerca de cada una. En la obra de Anderson y Anderson se encuentra un estudio más profundo de estas técnicas y sobre la forma de interpretarlas.

Escalas

Una escala es un conjunto de valores numéricos asignados a sujetos, objetos o comportamientos con el propósito de cuantificar y medir sus cualidades. Las escalas sirven para medir actitudes, valores e intereses. En contraposición con las pruebas, los resultados de estos instrumentos no indican éxito o fracaso, fuerza o debilidad; se limitan a medir el grado en que un individuo posee la característica. Por ejemplo, con una escala puede medirse la actitud de los universitarios hacia la religión.

La elaboración de escalas para medir actitudes, valores e intereses puede exigir la aplicación de diversas técnicas. En la siguiente exposición intentaremos presentar algunas de ellas.

Escalas de actitudes

Hay cuatro tipos principales de esas escalas: 1) escalas de evaluación, sumaria (escalas de Likert), 2) escalas de intervalos de aparición constante (escalas de Thurstone), 3) escalas acumulativas (escalas de Guttman), 4) escalas de diferencial semántico.

Escalas de Likert (método de evaluaciones sumarias). Presentan un número de enunciados negativos y positivos acerca de un objeto de actitud. Al responder a los puntos de estas escalas los sujetos indican si están firmemente de acuerdo, si están de acuerdo, indecisos, en desacuerdo, o en desacuerdo total con cada enunciado. El valor numérico que se asigna a cada respuesta depende del grado de acuerdo o desacuerdo con un enunciado individual.

La puntuación de un sujeto se determina sumando los valores asignados a las respuestas. Por ejemplo, se puede calificar una escala de Likert adscribiendo un valor de 2 puntos a cada respuesta que indique gran acuerdo con enunciados favorables, un valor de 1 punto al simple acuerdo, o a la indecisión, –1 al desacuerdo y –2 al desacuerdo total. En los enunciados desfavorables se invierte el procedimiento, ya que el desacuerdo se considera psicológicamente equivalente al acuerdo con un enunciado favorable.

No importa si 2 representa una puntuación alta y -2 una puntuación baja, o viceversa. Lo esencial es calificar las respuestas uniformemente, según la actitud que indiquen. Por supuesto el hecho de la actitud favorable corresponda a la *aprobación completa* o a la *desaprobación total* dependerá del contenido del enunciado.

Para construir una escala de Likert suelen seguirse los siguientes pasos:

1. Recopilar gran número de enunciados favorables y desfavorables sobre un objeto de actitud.
2. Seleccionar de entre ellos un número aproximadamente igual de enunciados favorables y desfavorables.
3. Administrar estos reactivos a varios individuos, pidiéndoles que indiquen sus opiniones acerca de cada uno y que digan si están en acuerdo total, de acuerdo, indecisos, en desacuerdo o en completo desacuerdo.
4. Calcular la puntuación de cada uno por medio del procedimiento descrito con anterioridad.
5. Analizar los reactivos para seleccionar aquellos que produzcan la mejor diferenciación. Mediante este estudio se descubre la correlación existente entre las puntuaciones totales de los sujetos y sus respuestas a cada reactivo.

Examinemos estos ejemplos sacados de los reactivos sobre “actitudes hacia el negro” que Likert incluyó en su “Encuesta de opiniones”.

Si se requiere la misma preparación el maestro de raza negra deberá percibir el mismo salario que el banco.	SA (2)	A (1)	U (0)	D (-1)	SD (-2)
Los barrios de negros deberán estar Separados de los blancos.	SA (-2)	A (-1)	U (0)	D (1)	SD (2)

Escalas de Thurstone

Método de los intervalos de aparición constante. Thurstone ideó un método para asignar valores escalares específicos a los reactivos que representan diferentes grados de actitud favorable. En la elaboración de este tipo de escala se siguen los pasos siguientes:

1. Recopilar gran número de enunciados sobre el objeto de actitud.
2. Presentarlos a cierto número de personas que les juzgarán. Suelen usarse de 50 a 100 jueces para evaluarlas. Trabajan de forma independiente y dividen todos los enunciados en siete, nueve u once categorías, conforme al grado favorable expresado. En el primer grupo se colocan los que consideren más favorables al objeto; en el segundo grupo los menos favorables, y así sucesivamente. El sexto grupo representa la posición neutral y el decimoprimero contiene los enunciaos más desfavorables. Es importante observar que esta clasificación no tiene nada que ver con las actitudes de los jueces hacia el objeto, sino que sólo representa sus juicios acerca de lo favorable de los enunciados.
3. Encontrar la escala que habrá de asignarse al valor de cada enunciado, calculando para ello la mediana de los pesos o posiciones que los jueces les asignaron en la escala. Se suprimen los reactivos que muestren gran variabilidad.

4 Seleccionar de 20 a 30 enunciados que tengan igual grado de dispersión en la escala. Tales enunciados vienen a constituir la escala de actitudes. Los siguientes reactivos con sus valores escalares provienen de la escala de Thurstone, que mide actitudes ante la iglesia.

Valor escalar

0.2 Creo que hoy la iglesia es la institución más importante de Estados Unidos.
1.5 Pienso que pertenecer a la iglesia es esencial para vivir en plenitud.
2.3 La liturgia me procura una sensación de tranquilidad y me alienta.
3.3 Me gusta la iglesia porque encuentro en ella un ambiente amoroso.
4.5 Creo en las enseñanzas de la iglesia, con algunas reservas.
5.6 Algunas veces pienso que la iglesia y la religión son necesarias pero en otras ocasiones lo dudo.
6.7 Creo en la sinceridad y en la bondad sin necesidad de asistir a ceremonias religiosas.
7.4 Me parece que la iglesia esta perdiendo terreno conforme avanza la educación.
8.3 Considero que las enseñanzas de la iglesia son demasiado superficiales y que por lo mismo tienen poca trascendencia social.
9.6 Pienso que la iglesia es un estorbo para la religión ya que todavía se basa en la magia, la superstición y el mito.
11.0 Creo que la iglesia es un parásito de la sociedad.

Para administrar una escala de Thurstone se pide a los sujetos que seleccionen de la lista los enunciados representativos de sus opiniones, o que escojan tres de ellas que se acerquen lo más posible a su posición al respecto. Por supuesto que los valores de la escala no vienen indicados en la escala de actitud, y los incisos se presentan en orden aleatorio. La puntuación de cada sujeto sería la media de los valores de las declaraciones que haya seleccionado.

Escalas de Guttman (técnica acumulativa)

Los críticos de las escalas de actitudes de Thurstone y Likert recalcan que contienen enunciados heterogéneos acerca de varias dimensiones de un objeto de actitud. Por ejemplo, en la escala de Thurstone que medía las actitudes hacia la guerra se separaron los enunciados éticos de los que hacían referencia a sus consecuencias económicas ni de los que reflejaban otros aspectos de las actitudes hacia ella. Esta combinación de diversas dimensiones sobre una escala puede hacer muy difícil interpretar con claridad las puntuaciones.

Guttman ideó una técnica para resolver este problema. Su método caracterizado como una escala unidimensional, pretende determinar si la actitud que se va a estudiar comprende en realidad una sola dimensión. Una actitud se considera unidimensional *sólo* si produce una escala acumulativa, aquella en que los enunciados se relacionen entre sí, de modo que un sujeto que está de acuerdo con el reactivo 2 lo estará también con el 1; el que coincida con el 3 coincidirá con el 1 y el 2, etc. Por lo tanto, los que le den su aprobación a un punto particular de este tipo de escala tendrán una puntuación más alta en la escala total que los que no estén de acuerdo con él.

Por ejemplo, consideremos los siguientes enunciados con que los sujetos deberán de estar de acuerdo o discrepar:

Puntuación	Concuerda con el reactivo			No concuerda con el reactivo		
	3	2	1	3	2	1
3	X	X	X	0	0	0
2	0	X	X	X	0	0
1	0	0	X	X	X	0
0				X	X	X

* El sujeto obtiene un punto cada vez que manifiesta estar de acuerdo

1. La asociación de padres y maestros es una institución que da buenos nos frutos.
2. La asociación de padres y maestros influye notablemente en el mejoramiento de las escuelas.
3. La asociación de padres y maestros constituye la organización más importante de Estados Unidos para el mejoramiento de las escuelas.

Figura 3

Si lo anterior es una escala acumulativa, deberá ser posible disponer todas las respuestas de los sujetos dentro de la estructura que aparece en el cuadro 3. Por tanto, si conocemos una puntuación individual podría decirse exactamente cuáles enunciados aprobó el sujeto. Por ejemplo, todos los individuos con una puntuación de 2 creen que la asociación de padres y maestros da buenos frutos y que influye notablemente en el mejoramiento de las escuelas, pero *no* piensan que sea la organización más importante en ese aspecto. Los sujetos pueden clasificarse según sus respuestas en la escala.

Al construir una escala acumulativa, ante todo ha de averiguarse si los reactivos forman o no una escala unidimensional. Para ello se analiza la reproductividad de las respuestas esto es, la proporción de las que caen dentro de una plantilla como la que se muestra en el cuadro 3. Tomando como base la puntuación total, se predice la estructura de respuestas a incisos particulares.

A continuación se estudia la estructura real de las respuestas y se mide el grado en que fueron reproducidos en la puntuación total. Una técnica consiste en dividir el número total de errores entre el número de respuestas y restar de uno el cociente. Guttman sugiere 0,90 como el mínimo coeficiente de reproductividad necesario para admitir que una serie de reactivos constituye una escala unidimensional o acumulativa.

Escalas de diferencial semántico.

El diferencial semántico es otra forma de medir las actitudes hacia objetos, sujetos o eventos. Esta técnica fue inventada y usada por Osgood, Suci y Tannenbaum. Se basa en la idea de que los objetos tienen dos tipos de significado para los individuos, el *denominativo* y el *connotativo,* que pueden clasificarse independientemente. Es fácil expresar el primero, no así el segundo.

Es factible y útil medir indirectamente el significado connotativo de los objetos mediante adjetivos bipolares y solicitando a los sujetos que para clasificar los objetos los comparen con esos adjetivos. Así, el significado que tenga un objeto para un individuo será la estructura de sus estimaciones del mismo sobre las escalas de adjetivos bipolares que hayan sido usadas. Osgood y sus colegas emplean escalas de siete puntos, con un punto medio igual a cero y de +3 a –3 para clasificaciones de actitud. A continuación presentamos algunos ejemplos:

Al obtener las clasificaciones que los individuos hacen de un objeto, el investigador puede determinar si la actitud de ellos hacia el objeto es positiva o negativa. Las puntuaciones de actitud logradas por cada respondiente se comparan con la actitud típica que hacia el objeto externa un grupo designado. Una puntuación de la actitud también se obtiene comparando las actitudes de varias personas hacia el objeto y la estructura de sus clasificaciones con las de los otros.

Bueno	+3	+2	+1	0	-1	-2	-3	Malo
Limpio	+3	+2	+1	0	-1	-2	-3	Sucio
Dulce	+3	+2	+1	0	-1	-2	-3	Amargo
Fuerte	+3	+2	+1	0	-1	-2	-3	Débil
Grande	+3	+2	+1	0	-1	-2	-3	Pequeño
Pesado	+3	+2	+1	0	-1	-2	-3	Ligero
Activo	+3	+2	+1	0	-1	-2	-3	Pasivo
Rápido	+3	+2	+1	0	-1	-2	-3	Lento
Caliente	+3	+2	+1	0	-1	-2	-3	Frío

A través de estudios de análisis factoriales, Osgood, Suci y Tannenbaum descubrieron tres grupos de adjetivos: *valorativos,* que corresponden a adjetivos como bueno y malo, o limpio y sucio; *de potencia,* que incluyen adjetivos como fuerte y débil, o grande y pequeño; *y de actividad,* que abarca adjetivos como activo y pasivo, rápido y lento. Del grupo, el valorativo es el más sobresaliente de los tres.

Los informes sobre la validez y confiabilidad de las escalas de diferencial semántico son generalmente satisfactorios. Los estudios de validez muestran coeficientes de correlación de un 0,80 entre las estimaciones derivadas del diferencial semántico y las de las escalas de Thurstone, Likert y Guttman.

Su confiabilidad de preprueba y posprueba viene a ser aproximadamente de 0,90, un resultado que es satisfactorio. Aunque, según parece, el diferencial semántico se utiliza menos que los otros tres tipos de escalas de actitudes es una técnica útil para medir las actitudes hacia los objetos.

Técnicas sociométricas

Sirven para estudiar la organización de los grupos sociales. El procedimiento básico, que puede modificarse de distintas maneras, consiste en solicitar a los miembros de un grupo particular que indiquen sus primeras, segundas y subsecuentes elecciones de compañeros según un criterio específico, por lo general en relación con una actividad concreta. Por ejemplo, ¿con quién le gustaría trabajar en este proyecto, junto a quién le gustaría sentarse? ¿con quien almorzaría o iría a jugar al salir de la escuela? El método sociométrico es en esencia un estudio de las selecciones hechas por los integrantes de un grupo.

SOCIOGRAMA DE UN GRUPO

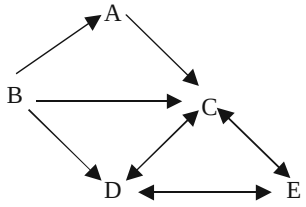


Figura 4

Las preferencias se trazan en un Sociograma, que muestra el patrón de las relaciones interpersonales. En la figura 4 aparece un Sociograma donde C es el miembro que acapara las preferencias y al que podría llamársele la "estrella". Obsérvese que D, C y E se escogen mutuamente. Esto representa una camarilla; es decir, tres o más individuos que muestran una preferencia mutua. B escoge a otros miembros y en cambio nadie escoge a él; es un *aislado.* Las preferencias que revele un Sociograma pueden cuantificarse y utilizarse en la investigación.

Los métodos sociométricos se aplican ampliamente en la psicología social y también en la investigación educacional, en la cual la posición sociométrica puede examinarse según su relación con otras variables como habilidad mental aprovechamiento, las preferencias de los maestros por los alumnos.

Observación directa

La observación directa y sistemática del comportamiento constituye en muchos casos el método de medición más adecuado. El investigador escoge la conducta que le interesa y elabora un procedimiento sistemático para identificarla, clasificarla y registrarla en una situación natural o preparada.

Un ejemplo excelente de la aplicación de este procedimiento en una situación natural es el estudio de Urban "Cambios del comportamiento resultantes de un estudio sobre enfermedades contagiosas". Los observadores registraron el número de casos de comportamiento indeseable (v. gr.: meter los dedos u otros objetos en la boca) y el número de conductas deseables; utilizar el pañuelo al toser o estornudar. Después de esto seleccionaron un grupo experimental al cual se impartió un curso de seis semanas sobre enfermedades contagiosas, diseñado para cambiar su comportamiento total y proporcionar información objetiva y conocimientos. Al final del curso reaparecieron los comportamientos indeseables. Se descubrió que éstos habían disminuido grandemente y que los deseables habían aumentado muchísimo en el grupo experimental, mientras que el cambio era mínimo en el grupo de control que no recibió el curso. Las observaciones realizadas doce semanas después demostraron que persistían las diferencias entre ambos grupos.

Hartshorne, May y Shuttleworth se valieron de la observación directa para medir rasgos como dominio de sí mismo, cooperatividad, veracidad y honestidad. Hicieron observaciones sobre niños ocupados en actividades escolares ordinarias y también prepararon situaciones artificiales para examinar conductas específicas. Por ejemplo, les administraron pruebas de vocabulario y de lectura, las recogieron y, sin que los niños lo supieran sacaron copias de las respuestas. Más tarde les dieron claves de respuesta y les pidieron que calificaran sus exámenes. La diferencia entre las puntuaciones que dieron los niños y las puntuaciones verdaderas obtenidas al calificar las copias de los exámenes proporcionó una medición del engaño.

Toda la observación directa ha de estar precedida de cinco pasos preliminares:

1. *Debe seleccionarse el aspecto del comportamiento que se va a observar.* Como no es posible observar todo lo que ocurre, el investigador tiene que escoger de antemano los comportamientos que va registrar y los que ignorará.
2. *Los comportamientos que pertenezcan a la categoría elegida tienen que estar claramente definidos.* Los observadores deben entender las acciones que clasificarán como, por ejemplo, conducta cooperativa o egoísta.
3. *Las personas que realizarán las observaciones deben haber recibido adiestramiento.* El adiestramiento y la práctica son necesarios para que el investigador pueda confiar en que seguirán el procedimiento establecido para observar, interpretar y comunicar las observaciones.
4. *Debe idearse un sistema para cuantificar las observaciones.* El investigador tiene que escoger un método corriente para contar los comportamientos observados. Por ejemplo, es necesario establecer de antemano si una acción y la reacción correspondiente serán consideradas como uno o como dos casos de la conducta. Un buen método es dividir el periodo de observación en breves periodos y registrar en cada uno, digamos de 30 segundos, si el sujeto mostró el comportamiento o no.
5. *Tienen que idearse procedimientos minuciosos para registrar el comportamiento.* No puede confiarse demasiado en la memoria de los observadores. La mejor solución es un sistema codificador que permita registrar inmediatamente, con una simple letra o un dígito, lo que se está observando. Se prefiere este sistema al narrativo, porque éste requiere mucho tiempo y atención por parte de los observadores.

Un ejemplo muy usado del plan codificador es el sistema Flanders para clasificar la conducta verbal en el aula. Tal como puede verse en la figura 5.

Este sistema proporciona categorías minuciosas y mutuamente exclusivas, cada una de las cuales puede registrarse como un solo digito. Los observadores expertos están capacitados para registrar un dígito cada tres segundos. La cadena de dígitos que aparezca podrá analizarse con facilidad para obtener no sólo un registro de las proporciones de la conducta verbal que caen dentro de cada categoría, sino también un cuadro de las que precedían o seguían a otras.

CATEGORÍAS DEL ANÁLISIS DE LA INTERACCIÓN SEGÚN FLANDERS

LENGUAJE DE MAESTRO:

	1.* *Acepta los sentimientos:* acepta y esclarece una actitud al tono emotivo de una alumna sin adoptar un gesto amenazador. Los sentimientos pueden ser positivos o negativos. Se incluyen sentimientos de predicción y de recuerdo.
	2.* *Alienta o elogia:* alienta o alaba la acción o el comportamiento de los alumnos. Rompe la tensión contando chistes, pero sin burlarse de otros; inclina la cabeza en señal de aprobaci6n o dice "ajá" o "prosiga".
Respuesta	3.* *Acepta o utiliza las ideas de los alumnos:* clarifica, construye o desarrolla las ideas que sugiere un alumno. Se incluyen las ampliaciones que haga de las ideas de los estudiantes; pero conforme aporte más de las propias éstas se desplazan a la categoría 5.
	4.* *Formula preguntas:* plantea una pregunta sobre el contenido a el procedimiento, basado en sus conocimientos pero con la intención de que el alumno la conteste.
	5.* *Instruye:* ofrece hechos u opiniones sobre contenidos o procedimientos; expresa sus ideas personales, da su propia explicación o cita a una autoridad en la materia que no sea una alumna.
	6.* *Imparte instrucciones:* da instrucciones, mandatos u ordena lo que el alumno debe hacer.
Iniciación	7.* Critica o justifica la autoridad: hace afirmaciones con el prop6sito de cambiar el comportamiento del alumno y lograr que adopte una conducta aceptable; regaña a veces; expone los motivos de sus acciones: demasiadas alusiones a sí mismo.
LENGUAJE DEL ALUMNO:	
Respuesta	8.* *Respuesta-lenguaje del alumno: conversación de los* alumnos en respuesta al maestro. Este inicia el contacto, les pide que hablen o estructura la situación. La libertad para expresar las propias ideas está limitada.
Iniciación	9.* *Iniciación-lengua del alumno:* conversación de los alumnos iniciada por ellos mismos. Expresan sus propias ideas; comienzan un nuevo tema: se les permite exponer opiniones y seguir una línea de pensamiento, les gusta formular preguntas profundas; van más allá de la estructura actual.
Silencio	10.* *Silencio o confusión:* pausas, cortos periodos de silencio y periodos de confusión en los cuales el observador no entenderá la comunicación.

* Estos números no implican ninguna escala. Cada número es de índole clasificatoria y designa un tipo especial de proceso de comunicación. Al escribir estos números durante la observación se pretende enumerar, y no juzgar, una posición en la escala.

FUENTE: Ned A. Flanders: Analyzing Teaching Behavior (Boston: Addison-Wesley, 1970).

Figura 5

Resumen

Una tarea importante de los investigadores de las ciencias del comportamiento consiste en seleccionar los instrumentos de medición para cuantificar la información. A los sujetos se les formulan preguntas directas mediante las entrevistas o los cuestionarios. Ambas técnicas pueden ser estructuradas o inestructuradas. En las entrevistas y cuestionarios inestructurados se presentan las preguntas a los sujetos y se registran sus respuestas libres. En las formas estructuradas las preguntas están organizadas y a los sujetos se les ofrece una selección limitada de respuestas.

Las pruebas constituyen los instrumentos más importantes para la recopilación de datos en la investigación educacional. Una prueba es una serie de estímulos que producen el rendimiento típico en el sujeto. Las Pruebas de aprovechamiento son excelentes ejemplos de esta clase de medición. Hay una gran variedad de ellas que proporcionan normas que pueden emplearse se como base de comparación. Las pruebas de inteligencia son instrumentos para evaluar las capacidades verbales y no verbales de un individuo. Los inventarios de la personalidad tienen por objeto medir las características personales del sujeto.

Las escalas de actitudes son instrumentos para medir las creencias, los sentimientos y reacciones del individuo a ciertos objetos. Los tipos más importantes son las escalas de Likert, las de Thurstone, las de Guttman y el Diferencial Semántico.

Las técnicas sociométricas sirven para estimar la posición de un individuo entre sus compañeros. Permiten identificar a los miembros populares de los grupos (las estrellas), a los aislados y a las camarillas.

Los métodos de observación directa fueron creados para observar sistemáticamente el comportamiento.

RECOLECCIÓN DE LOS DATOS

Roberto Hernández Sampieri*

¿QUÉ IMPLICA LA ETAPA DE RECOLECCIÓN DE LOS DATOS?

Una vez que seleccionamos el diseño de investigación apropiado y la muestra adecuada de acuerdo con nuestro problema de estudio e hipótesis, la siguiente etapa consiste en recolectar los datos pertinentes sobre las variables involucradas en la investigación.

Recolectar los datos implica tres actividades estrechamente vinculadas entre sí:

a) *Seleccionar un instrumento de medición de los disponibles en el estudio del comportamiento* o desarrollar uno (el instrumento de recolección de los datos). Este instrumento debe ser válido y confiable, de lo contrario no podemos basarnos en sus resultados.
b) *Aplicar ese instrumento de medición*. Es decir, obtener las observaciones y mediciones de las variables que son de interés para nuestro estudio (medir variables).
c) *Preparar las mediciones obtenidas* para que puedan analizarse correctamente (a esta actividad se le denomina codificación de los datos).

¿QUÉ SIGNIFICA MEDIR?

De acuerdo con la definición clásica del termino –ampliamente difundida– medir significa "asignar números a objetos y eventos de acuerdo a reglas" (Stevens, 1951). Sin embargo, como señalan Carmines y Zeller (1979), esta definición es más apropiada para las ciencias físicas que para las ciencias sociales, ya que varios de los fenómenos que son medidos en éstas no pueden caracterizarse como objetos o eventos, puesto que son demasiado abstractos para ello. La disonancia cognitiva, la alienación, el producto nacional bruto y la credibilidad son conceptos tan abstractos para ser considerados "cosas que pueden verse o tocarse" (definición de objeto) o solamente como "resultado, consecuencia o producto" (definición de evento) (Carmines y Zeller, 1979, p. 10).

---

* Tomado de: HERNÁNDEZ, Roberto: *Metodología de la investigación.* México, D.F., Mc Graw-Hill. 1998. pp. 241-346.

Este razonamiento nos hace sugerir que es más adecuado definir la medición como "el proceso de vincular conceptos abstractos con indicadores empíricos", proceso que se realiza mediante un plan explícito y organizado para clasificar (y frecuentemente cuantificar) los datos disponibles –los indicadores– en términos del concepto que el investigador tiene en mente (Carmines y Zeller, 1979, p. 10). Y en este proceso, el instrumento de medición o de recolección de los datos juega un papel central. Sin él no hay observaciones clasificadas.

La definición sugerida incluye dos consideraciones: La primera es desde el punto de vista empírico y se resume en que el centro de atención es la respuesta observable (sea una alternativa de respuesta marcada en un cuestionario, una conducta grabada vía observación o una respuesta dada a un entrevistador). La segunda es desde una perspectiva teórica y se refiere a que el interés se sitúa en el concepto subyacente no observable que es representado por la respuesta (Carmines y Zeller, 1979). Así, los registros del instrumento de medición representan valores observables de conceptos abstractos. Un *instrumento de medición adecuado es aquel que registra datos observables que representan verdaderamente a los conceptos o variables que el investigador tiene en mente*.

En toda investigación aplicamos un instrumento para medir las variables contenidas en las hipótesis (y cuando no hay hipótesis, simplemente para medir las variables de interés). Esa medición es efectiva cuando el instrumento de recolección de los datos realmente representa a las variables que tenemos en mente. Si no es así nuestra medición es deficiente y por lo tanto la investigación no es digna de tomarse en cuenta. Desde luego, no hay medición perfecta, es prácticamente imposible que representemos fielmente variables tales como la inteligencia, la motivación, el nivel socioeconómico, el liderazgo democrático, la actitud hacia el sexo y otras más; pero sí debemos de acercarnos lo más posible a la representación fiel de las variables a observar, mediante el instrumento de medición que desarrollemos.

¿QUE REQUISITOS DEBE CUBRIR UN INSTRUMENTO DE MEDICIÓN?

Toda medición o instrumento de recolección de los datos debe reunir dos requisitos esenciales: *confiabilidad* y *validez*. La confiabilidad de un instrumento de medición se refiere al grado en que su aplicación repetida al mismo sujeto u objeto, produce iguales resultados. Por ejemplo, si yo midiera en este momento la temperatura ambiental mediante un termómetro y me indicara que hay 22°C. Un minuto más tarde consultara otra vez y el termómetro me indicara que hay 5°C. Tres minutos después observara el termómetro y ahora me indicara que hay 40°C.

Este termómetro no sería confiable (su aplicación repetida produce resultados distintos). Igualmente, si una prueba de inteligencia la aplico hoy a un grupo de personas y me proporciona ciertos valores de inteligencia; la aplico un mes después y me proporciona valores diferentes, al igual que en subsecuentes mediciones. Esa prueba no es confiable (analícense los valores de la figura 1, suponiendo que los coeficientes de inteligencia puedan oscilar entre 95 y 150). Los resultados no son consistentes; no se puede "confiar" en ellos.

EJEMPLO DE RESULTADOS PROPORCIONADOS POR UN INSTRUMENTO DE MEDICIÓN SIN CONFIABILIDAD

PRIMERA APLICACIÓN		SEGUNDA APLICACIÓN		TERCERA APLICACIÓN	
Martha	130	Laura	131	Luis	140
Laura	125	Luis	130	Teresa	129
Arturo	118	Marco	127	Martha	124
Luís	112	Arturo	120	Rosa Maria	120
Marco	110	Chester	118	Laura	109
Rosa Maria	110	Teresa	118	Chester	108
Chester	108	Martha	115	Arturo	103
Teresa	107	Rosa Maria	107	Marco	101

Figura 1

La confiabilidad de un instrumento de medición se determina mediante diversas técnicas, las cuales se comentarán brevemente después de revisar el concepto de validez.

La validez, en términos generales, se refiere al grado en que un instrumento realmente mide la variable que pretende medir. Por ejemplo, un instrumento válido para medir la inteligencia debe medir la inteligencia y no la memoria. Una prueba sobre conocimientos de Historia debe medir esto y no conocimientos de literatura histórica. Aparentemente es sencillo lograr la validez. Después de todo –como dijo un estudiante– "pensamos en la variable y vemos cómo hacer preguntas sobre esa variable". Esto sería factible en unos cuantos casos (como lo sería el "sexo" de una persona). Sin embargo, la situación no es tan simple cuando se trata de variables como la motivación, la calidad de servicio a los clientes, la actitud hacia un candidato político y menos aún con sentimientos y emociones, así como diversas variables con las que trabajamos en ciencias sociales. La validez es una cuestión más compleja que debe alcanzarse en todo instrumento de medición que se aplica. Kerlinger (1979, p. 138) plantea la siguiente pregunta respecto a la validez: ¿Está usted midiendo lo que usted cree que esta midiendo? Si es así, su medida es válida; si no, no lo es.

La validez es un concepto del cual pueden tenerse diferentes tipos de evidencia (Wiersma, 1986; Gronlund, 1985): 1) *evidencia relacionada con el contenido*, 2) *evidencia relacionada con el criterio y* 3) *evidencia relacionada con el constructo*. Hablemos de cada una de ellas.

1) Evidencia relacionada con el contenido

La validez de contenido se refiere al grado en que un instrumento refleja un dominio específico de contenido de lo que se mide. Es el grado en que la medición representa al concepto medido (Bohrnstedt, 1976). Por ejemplo, una prueba de operaciones aritméticas no tendrá validez de contenido si incluye sólo problema de resta y excluye problema de suma, multiplicación o división (Carmines y Zeller, 1979). O bien, una prueba de conocimientos sobre las canciones de "Los Beatles" no deberá basarse solamente en sus álbumes "Let it Be" y "Abbey Road", sino que debe incluir canciones de todos sus discos.

Un instrumento de medición debe *contener* representados a todos los ítems del dominio de contenido de las variables a medir. Este hecho se ilustra en la figura 2.

2) Evidencia relacionada con el criterio

La validez de criterio establece la validez de un instrumento de medición comparándola con algún criterio extremo. Este criterio es un estándar con el que se juzga la validez del instrumento (Wiersma, 1986). Entre los resultados del instrumento de medición se relacionen más al criterio, la validez del criterio será mayor. Por ejemplo, un investigador valida un examen sobre manejo de aviones, mostrando la exactitud con que el examen predice qué tan bien un grupo de pilotos puede operar un aeroplano.

Si el criterio se fija en el *presente*, se habla de *validez concurrente* (los resultados del instrumento se correlacionan con el criterio en el mismo momento o punto del tiempo). Por ejemplo, un cuestionario para detectar las preferencias del electorado por los distintos partidos contendientes, puede validarse aplicándolo tres o cuatro días antes de la elección y sus resultados compararlos con los resultados finales de la elección (si no hay fraude –desde luego–).

Si el criterio se fija en el *futuro* se habla de *validez predictiva*. Por ejemplo, una prueba para determinar la capacidad administrativa de altos ejecutivos se puede validar comparando sus resultados con el futuro desempeño de los ejecutivos medidos.

ILUSTRACIÓN DE UN INSTRUMENTO DE MEDICIÓN CON VALIDEZ DE CONTENIDO VERSUS CON UNO QUE CARECE DE ÉSTA

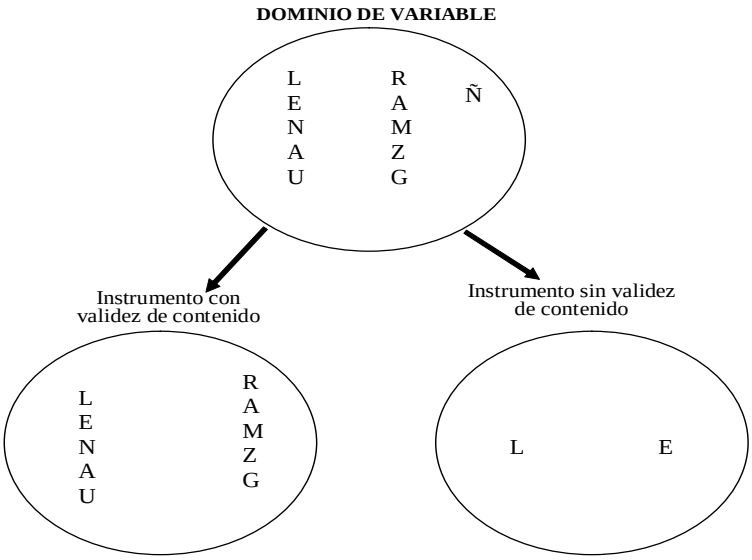


Figura 2

3) *Evidencia relacionada con el constructo*

La validez de constructo es probablemente la más importante sobre todo desde una perspectiva científica y se refiere al grado en que una medición se relaciona consistentemente con otras mediciones de acuerdo con hipótesis derivadas teóricamente y que conciernen a los conceptos (o constructos) que están siendo medidos. Un constructo es una variable medida y que tiene lugar dentro de una teoría o esquema teórico.

Por ejemplo, supongamos que un investigador desea evaluar la validez de constructo de una medición particular, digamos una escala de motivación intrínseca: "el Cuestionario de Reacción a Tareas", versión mexicana (Hernández–Sampieri y Cortes, 1982). Estos autores sostienen que el nivel de motivación intrínseca hacia una tarea está relacionado positivamente con el grado de persistencia adicional en el desarrollo de la tarea (v.g., los empleados con mayor motivación intrínseca son los que suelen quedarse más tiempo adicional una vez que concluye su jornada).

Consecuentemente, la predicción teórica es que a mayor motivación intrínseca, mayor persistencia adicional en la tarea. El investigador administra dicho cuestionario de motivación intrínseca a un grupo de trabajadores y también determina su persistencia adicional en el trabajo. Ambas mediciones son correlacionadas. Si la

correlación es positiva y sustancial, se aporta evidencia para la validez de constructo del Cuestionario de Reacción a Tareas, versión mexicana (a la validez para medir la motivación intrínseca).

La validez de constructo incluye tres etapas:

1) Se establece y específica la relación teórica entre los conceptos (sobre la base del marco teórico).
2) Se correlacionan ambos conceptos y se analiza cuidadosamente la correlación.
3) Se interpreta la evidencia empírica de acuerdo a qué tanto clarifica la validez de constructo de una medición en particular.

El proceso de *validación de un constructo* está vinculado con la teoría. No es posible llevar a cabo la validación de constructo, a menos que exista un marco teórico que soporte a la variable en relación con otras variables. Desde luego, no es necesaria una teoría sumamente desarrollada, pero sí investigaciones que hayan demostrado que los conceptos están relacionados. Entre más elaborado y comprobado se encuentre el marco teórico que apoya la hipótesis, la validación de constructo puede arrojar mayor luz sobre la validez de un instrumento de medición. Y mayor confianza tenemos en la validez de constructo de una medición, cuando sus resultados se correlacionan significativamente con un mayor número de mediciones de variables que teóricamente y de acuerdo con estudios antecedentes están relacionadas. Esto se representa en la figura 3.

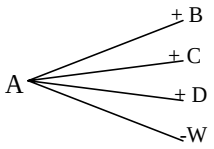
GRAFICACIÓN DE UN INSTRUMENTO CON VALIDEZ DE CONSTRUCTO

Instrumento mide constructo "A"

Teoría

(Investigaciones hechas) encontraron que A se relaciona positivamente con B, C y D. Y negativamente con W.

Si el instrumento mide realmente A, sus resultados deben relacionarse positivamente con los resultados obtenidos en las mediciones B, C y D y relacionarse negativamente con resultados de W.



El instrumento parece realmente medir "A"

Para analizar las posibles interpretaciones de evidencia negativa en la validez de constructo, se sugiere consultar a Cronbach y Meehl (1955) y Cronbach (1984).

VALIDEZ TOTAL = VALIDEZ DE CONTENIDO + VALIDEZ DE CRITERIO + VALIDEZ DE CONSTRUCTO

Figura 3

Así, la validez de un instrumento de medición se evalúa sobre la base de tres tipos de evidencia. Entre mayor evidencia de validez de contenido, validez de criterio y validez de constructo tenga un instrumento de medición; éste se acerca más a representar la variable o variables que pretende medir.

Cabe agregar que *un instrumento de medición puede ser confiable pero no necesariamente válido* (un aparato –por ejemplo– puede ser consistente en los resultados que produce, pero no medir lo que pretende). Por ello es requisito que el instrumento de medición demuestre ser confiable y válido. De no ser así, los resultados de la investigación no los podemos tomar en serio.

FACTORES QUE PUEDEN AFECTAR LA CONFIABILIDAD Y VALIDEZ

Hay diversos factores que pueden afectar la confiabilidad y la validez de los instrumentos de medición.

El primero de ellos es la *improvisación*. Algunas personas creen que elegir un instrumento de medición o desarrollar uno es algo que puede tomarse a la ligera. Incluso algunos profesores piden a los alumnos que construyan instrumentos de medición de un día para otro, o lo que es casi lo mismo, de una semana a otra. Lo cual habla del poco o nulo conocimiento del proceso de elaboración de instrumentos de medición. Esta improvisación genera –casi siempre– instrumentos poco válidos o confiables y no debe existir en la investigación social (menos aún en ambientes académicos). Aun a los investigadores experimentados les toma tiempo desarrollar un instrumento de medición. Es por ello que los construyen con cuidado y frecuentemente están desarrollándolos, para que cuando los necesiten con premura se encuentren preparados para aplicarlos, pero no los improvisan. Además, para poder construir un instrumento de medición se requiere conocer muy bien a la variable que se pretende medir y la teoría que la sustenta. Por ejemplo, generar o simplemente seleccionar un instrumento que mida la inteligencia, la personalidad o los usos y gratificaciones de la televisión para el niño, requiere amplios conocimientos en la materia, estar actualizados al respecto y revisar cuidadosamente la literatura correspondiente.

El segundo factor es que a veces se *utilizan instrumentos desarrollados* en el extranjero que no han sido validados a nuestro contexto: cultura y tiempo. Traducir un instrumento –aun cuando adaptemos los términos a nuestro lenguaje y los contextualicemos– no es de ninguna manera (ni remotamente) validarlo. Es un primer y necesario paso, pero sólo es el principio. Por otra parte, hay instrumentos que fueron validados en nuestro contexto pero hace mucho tiempo. Hay instrumentos que hasta el lenguaje nos suena “arcaico”. Las culturas, los grupos y las personas

cambian; y esto debemos tomarlo en cuenta al elegir o desarrollar un instrumento de medición.

Un tercer factor es que en ocasiones el instrumento resulta inadecuado para las personas a las que se les aplica: no es empático. Utilizar un lenguaje muy elevado para el respondiente, no tomar en cuenta diferencias en cuanto a sexo, edad, conocimientos, capacidad de respuesta, memoria, nivel ocupacional y educativo, motivación para responder y otras diferencias en los respondientes; son errores que pueden afectar la validez y confiabilidad del instrumento de medición.

Un cuanto factor que puede influir está constituido por las condiciones en las que se aplica el instrumento de medición. Si hay ruido, hace mucho frío (por ejemplo en una encuesta de casa en casa), el *instrumento* es demasiado largo o tedioso, son cuestiones que pueden afectar negativamente la validez y la confiabilidad. Normalmente en los experimentos se puede contar con instrumentos de medición más largos y complejos que en los diseños no experimentales. Por ejemplo, en una encuesta pública sería muy difícil poder aplicar una prueba larga o compleja.

Por otra parte, *aspectos mecánicos* tales como que si el instrumento es escrito, no se lean bien las instrucciones, falten páginas, no haya espacio adecuado para contestar, no se comprendan las instrucciones, también pueden influir de manera negativa.

¿CÓMO SE SABE SI UN INSTRUMENTO DE MEDICIÓN ES CONFIABLE Y VÁLIDO?

En la práctica es casi imposible que una medición sea perfecta. *Generalmente se tiene un grado de error*. Desde luego, se trata de que este error sea el mínimo posible. Es por esto que la medición de cualquier fenómeno se conceptualiza con la siguiente fórmula básica:

$$X = t + e$$

Donde "X" representa los valores observados (resultados disponibles), "t" son los valores verdaderos y "e" *es e1 grado de error en la medición*. Si no hay error de medición ("e" es igual a cero), el valor observado y el verdadero son equivalentes. Esto puede verse claramente así:

$$X = t + 0$$

$$X = t$$

Esta situación representa el ideal de la medición. Entre mayor sea el error al medir, el valor que observamos (y que es en el que nos basamos) se aleja más del valor real o verdadero. Por ejemplo, si medimos la motivación de un individuo y esta

medición está contaminada por un grado de error considerable, la motivación registrada por el instrumento será bastante diferente de la motivación real que tiene ese individuo. Por ello es *importante que el error sea reducido lo más posible*. Pero ¿cómo sabemos el grado de error que tenemos en una medición? Calculando la confiabilidad y validez.

Cálculo de la confiabilidad

Existen diversos procedimientos para calcular la confiabilidad de un instrumento de medición. Todos utilizan fórmulas que producen coeficientes de confiabilidad. Estos coeficientes pueden oscilar entre 0 y 1. Donde un coeficiente de 0 significa nula confiabilidad y 1 representa un máximo de confiabilidad (confiabilidad total). Entre más se acerque el coeficiente a cero (0), hay mayor error en la medición. Esto se ilustra en la figura 4.

INTERPRETACIÓN DE UN COEFICIENTE DE CONFIABILIDAD

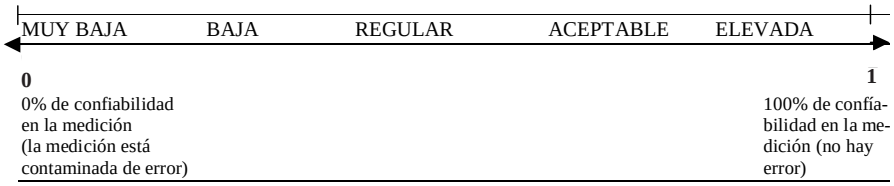


Figura 4

Los procedimientos más utilizados para determinar la confiabilidad mediante un coeficiente son:

1. *Medida de estabilidad* (confiabilidad por test–retest). En este procedimiento un mismo instrumento de medición (o ítems o indicadores) es aplicado dos o más veces a un mismo grupo de personas, después de un periodo de tiempo. Si la correlación entre los resultados de las diferentes aplicaciones es altamente positiva, el instrumento se considera confiable. Se trata de una especie de diseño panel. Desde luego, el periodo de tiempo entre las mediciones es un factor a considerar. Si el periodo es largo y la variable susceptible de cambios, ello puede confundir la interpretación del coeficiente de confiabilidad obtenido por este procedimiento, y si el periodo es corto las personas pueden recordar cómo contestaron en la primera aplicación del instrumento, para aparecer como más consistentes de lo que son en realidad (Bohrnstedt, 1976).
2. Método de formas alternativas o paralelas. En este procedimiento no se administra el mismo instrumento de medición, sino dos o más versiones

equivalentes de éste. Las versiones son similares en contenido, instrucciones, duración y otras características. Las versiones –generalmente dos– son administradas a un mismo grupo de personas dentro de un periodo de tiempo relativamente corto. El instrumento es confiable si la correlación entre los resultados de ambas administraciones es significativamente positiva. Los patrones de respuesta deben variar poco entre las aplicaciones.

3. Método de mitades partidas (split–halves). Los procedimientos anteriores (medida de estabilidad y método de formas alternas), requieren cuando menos dos administraciones de la medición en el mismo grupo de individuos. En cambio, el método de mitades partidas requiere sólo una aplicación de la medición. Específicamente, el conjunto total de ítems (o componentes) es dividido en dos mitades y las puntuaciones o resultados de ambas son comparados. Si el instrumento es confiable, las puntuaciones de ambas mitades deben estar fuertemente correlacionadas. Un individuo con baja puntuación en una mitad, tenderá a tener también una baja puntuación en la otra mitad. El procedimiento se diagrama en la figura 5.

ESQUEMA DEL PROCEDIMIENTO DE MITADES-PARTIDAS

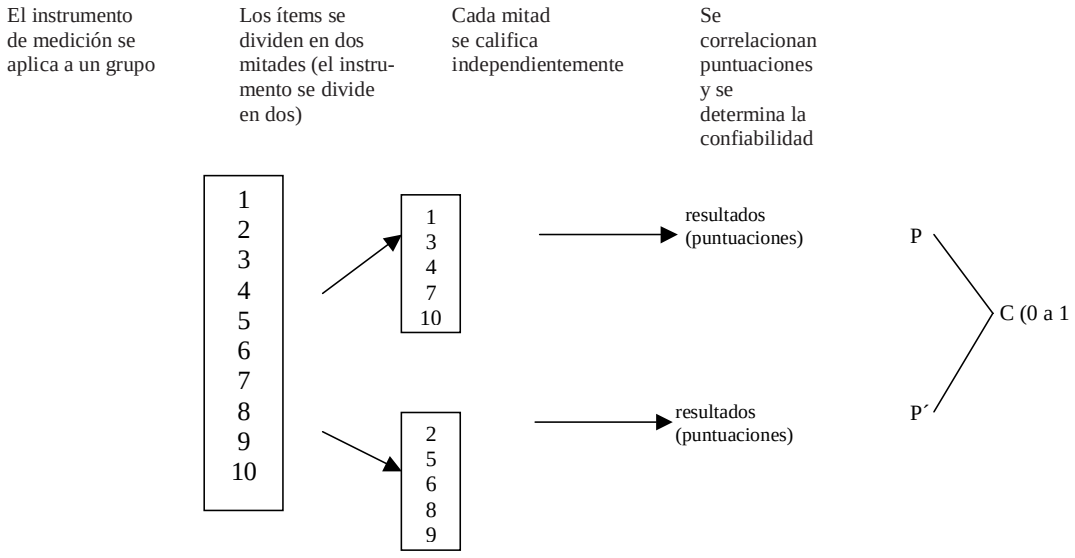


Figura 5

Al dividir los ítems, estos se empareja en contenido y dificultad

La confiabilidad varía de acuerdo al número de ítems que incluya el instrumento de medición. Cuantos más ítems la confiabilidad aumenta (desde luego, que se refieran a la misma variable). Esto resulta lógico, veámoslo con un ejemplo cotidiano: Si se desea probar qué tan confiable o consistente es la lealtad de un amigo hacia nuestra persona, cuantas más pruebas le pongamos, su confiabilidad será mayor. Claro está que demasiados ítems provocarán cansancio en el respondiente.

4. *Coeficiente alfa de Cronbach*. Este coeficiente desarrollado por J. L. Cronbach requiere una sola administración del instrumento de medición y produce valores que oscilan entre 0 y 1. Su ventaja reside en que no es necesario dividir en dos mitades a los ítems del instrumento de medición, simplemente se aplica la medición y se calcula el coeficiente.
5. *Coeficiente* KR–20. Kuder y Richardson (1937) desarrollaron un coeficiente para estimar la confiabilidad de una medición, su interpretación es la misma que la del coeficiente alfa.

Cálculo de la validez

La validez de contenido es compleja de obtener. Primero, es necesario revisar cómo ha sido utilizada la variable por otros investigadores. Y en base a dicha revisión elaborar un universo de ítems posibles para medir la variable y sus dimensiones (el universo tiene que ser lo más exhaustivo que sea factible). Posteriormente, se consulta con investigadores familiarizados con la variable para ver si el universo es exhaustivo. Se seleccionan los ítems bajo una cuidadosa evaluación. Y si la variable tiene diversas dimensiones o facetas que la componen, se extrae una muestra probabilística de ítems (ya sea al azar o estratificada –cada dimensión constituiría un estrato–). Se administran los ítems, se correlacionan las puntuaciones de los ítems entre sí (debe haber correlaciones altas, especialmente entre ítems que miden una misma dimensión) (Bohmstedt, 1976), y se hacen estimaciones estadísticas para ver si la muestra es representativa. Para calcular la validez de contenido son necesarios varios coeficientes.

La validez de criterio es más sencilla de estimar, lo único que hace el investigador es correlacionar su medición con el criterio, y este coeficiente es el que se toma como coeficiente de validez (Bohrnstedt, 1976). Esto podría representarse así:

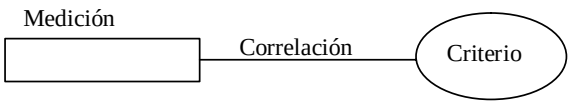


Figura 6

La validez de constructo se suele determinar mediante un procedimiento denominado "Análisis de Factores". Su aplicación requiere de sólidos conocimientos estadísticos y del uso de un programa estadístico apropiado en computadora. Para quien desee compenetrarse con esta técnica recomendamos consultar a Harman (1967), Gorsuch (1974), Nie et al. (1975), On–Kim y Mueller (1978 y 1978) y Hunter (1980). Asímismo, para aplicarlos se sugiere revisar a Nie et al. (1975), Cooper y Curtis (1976) y –en español– Padua (1979). Aunque es requisito conocer el programa estadístico para computadora.

CAPÍTULO II

EL MUESTREO

EL MUESTREO

Una vez que el investigador ha diseñado la estrategia para contrastar la hipótesis y ha elaborado sus instrumentos de acopio de datos, debe resolver el problema de identificar a los sujetos con los que trabajará o cuál será el objeto de estudio, para saber a quiénes o dónde aplicará los instrumentos, ya sea de medición, de observación o simplemente de acopio de información.

En esta etapa el investigador debe definir la unidad de análisis que, en el caso de la investigación de la conducta pueden ser personas, organizaciones, locales, métodos didácticos, etc. Una vez establecida la unidad de análisis, debe delimitar la población que será estudiada. Sobre esta población el investigador tratará de generalizar los resultados de su trabajo.

1. POBLACIÓN Y MUESTRA

Una **población** es la totalidad de sujetos o elementos que tienen características comunes. En otras palabras, una población es la totalidad de los miembros de la unidad de análisis. El concepto de población equivale al concepto de conjunto y éste es delimitado por el investigador según los criterios que considere pertinentes. Una población así conceptualizada será más grande o más pequeña, es decir, el tamaño de la población dependerá de la definición que el investigador formule. Si define la población como el total de personas que estudian educación superior en el Perú, el número de esta población es una cifra diferente a la cifra que corresponde al número de individuos varones que estudian educación superior en el Perú. También constituye una población el número de participantes de los programas de maestría de las universidades nacionales, este número existe y es diferente al número de esta otra población: el número de participantes de los programas de maestría en educación de la Universidad Nacional Mayor de San Marcos. Es muy importante que el investigador defina claramente la población porque, como se ha visto, de esta definición depende el número de los elementos del conjunto.

Según las exigencias de la estrategia para contrastar hipótesis, el investigador debe definir la población. Muchos investigadores no definen con claridad la población ni mucho menos el número de sus elementos y directamente identifican la muestra.

En este caso, los investigadores actúan a ciegas, pues no tienen claro cuál es la unidad de análisis con la que van a trabajar, ni tienen idea del tamaño de la población. Si por ejemplo se trata de hacer una investigación acerca de los hábitos de lectura de estudiantes universitarios, la unidad de análisis es *estudiantes universitarios*, pero es necesario definir qué estudiantes universitarios: ¿estudiantes universitarios de universidades nacionales o de universidades particulares?, ¿estudiantes universitarios de facultades de medicina o de ingeniería?, ¿estudiantes universitarios varones o mujeres? Las respuestas a estas interrogantes conducen a números diferentes de elementos del conjunto, pues el número de estudiantes universitarios del país es diferente del número de estudiantes universitarios de universidades nacionales. El número de estudiantes universitarios del Perú, no es el mismo que el número de estudiantes universitarios latinoamericanos. Es pues muy importante definir la población, porque de esta definición se deriva el número de sus elementos.

La población se llama también universo y su estudio se realiza mediante el **censo**, es decir, mediante el conteo, uno a uno, de todos los elementos del conjunto. Son ejemplos de poblaciones o universos, los alumnos que estudian educación superior en el Perú, los docentes sin título profesional de la ciudad de Lima o la totalidad de estudiantes de post grado de la Universidad Nacional Mayor de San Marcos.

La **muestra** es un sub conjunto de la población. Para que un sector de la población sea considerado como muestra es necesario que todos los elementos de ella pertenezcan a la población, por eso se dice que una muestra debe ser representativa de la población, es decir, debe tener las mismas características generales de la población. No se consideran muestras si algunos sujetos de la supuesta muestra no pertenecen a la población. El estudio de la muestra se realiza empleando técnicas de **muestreo**.

El **muestreo** es la técnica a través de la cual se estudia la muestra. Cuando se desea conocer, por ejemplo, la totalidad de habitantes del Perú, se necesita recurrir a un censo que es, como se ha dicho, el conteo uno a uno de todos los habitantes del país. El resultado del censo es una cifra exacta que se denomina **parámetro**. Pero es el caso que los censos, por la complejidad que demanda su ejecución y por los recursos económicos que requiere, no se realizan a cada momento. Por otra parte, los resultados de un censo no pueden ser conocidos de inmediato, por lo que la pregunta sobre la población del Perú queda sin respuesta inmediata. Pero como el investigador necesita una respuesta inmediata, recurre a otras formas de estimar el número de habitantes del Perú y, vía aproximaciones o análisis de los últimos datos censales y de los índices de mortalidad y natalidad en un período determinado, puede identificar una cifra, no exacta pero aproximada, que se denomina **estadígrafo**. Como resultado de la aplicación de la técnica de muestreo, el investigador obtiene

un estadígrafo, una cifra lograda por estimación, cálculo o aproximación al número real de los elementos de la población.

Por eso se dice que *el muestreo es una técnica que permite conocer los parámetros de la población a través de los estadígrafos de la muestra*. Esto quiere decir que la muestra, al ser considerada un sub conjunto de la población tiene que ser, necesariamente, representativa de la población para poder generalizar los datos que se obtengan al estudiar la muestra.

Una muestra no representativa es una muestra *sesgada*, por no poseer las mismas características de la población. Por ejemplo, si se quiere estudiar el índice de eficiencia de un Centro de Estudios Pre Universitarios, no se puede aplicar una prueba a todos los estudiantes que ingresaron en la Universidad, porque no todos los estudiantes que ingresaron provienen del mismo Centro de Estudios Pre Universitarios.

El conocimiento aproximado del número de elementos de la población es válido siempre y cuando se ubique dentro de los niveles de tolerancia o de variación de los datos. Por el contrario, cuando la discrepancia entre el estadígrafo y el parámetro es demasiada, no será posible conocer la real dimensión de la población. Si supuestamente se conociera el parámetro y se aceptara que el número de habitantes del Perú, el día de hoy, es de 28’867,563 habitantes, el estadígrafo que se obtendría sería válido si se aproximara a esta cifra ideal, como por ejemplo decir que la población del Perú, el día de hoy, se aproxima a los 29 millones de habitantes o se dijera que la población del Perú supera los 28 millones y medio de habitantes. Estas últimas son cifras calculadas, cifras estimadas, no son el resultado de un censo, son estadígrafos que tienen la particularidad de aproximarse al parámetro. Pero si dijéramos que la población del Perú supera los 100 millones de habitantes o se dijera que el número de habitantes del Perú es menos de 6 millones, entonces, no se estaría contribuyendo al conocimiento del parámetro a través de los estadígrafos, el investigador se estaría engañando con respecto al número de habitantes que hay, hoy día, en el Perú, porque estaría proporcionando cifras que no están próximas al parámetro y, con toda seguridad, para obtenerlas, no habría realizado los cálculos adecuados.

Para que un estadígrafo sea válido debe coincidir con el parámetro, con más o menos aproximación. En caso contrario, tal cifra no puede ser considerada un estadígrafo porque no se aproxima al parámetro.

La técnica del muestreo se usa cuando la población es relativamente grande y no se puede censar a todos los sujetos y hallar parámetros. Por el contrario, si la población de estudio es pequeña, no será necesario emplear la técnica del muestreo y será preferible disponer de datos censales o parámetros. Así, cuando la población es relativamente pequeña, la muestra tenderá a ser más grande. Y, por el contrario, si la población es demasiado grande, la muestra tenderá a ser relativamente pequeña.

La técnica de muestreo permite generalizar los datos obtenidos para todos los elementos de la población y no sólo para los elementos de la muestra. También se puede generalizar para otras poblaciones que tengan características similares a las que se estudia. Esta es la manera inferencial que hace posible la generalización del conocimiento y, por consiguiente, el avance de la ciencia.

Para obtener muestras, el investigador debe tener en cuenta lo siguiente:

La muestra debe ser una parte de la población fácilmente accesible. Si por ejemplo se desea hacer el análisis de la pureza del agua potable que se consume en Lima, los investigadores podrán abrir cualquier caño y llenar sus probetas de laboratorio con el agua que salga de dichos caños. Ahora si existiese alguna duda de que el agua que sale del caño no proviene de la planta central de procesamiento de agua, los investigadores pueden constituirse en dicha planta de tratamiento y llenar sus probetas en la laguna que contiene el agua lista para distribuirse en la ciudad.

Si un investigador desea estudiar la influencia que determinados *métodos didácticos* producen en el *aprendizaje* de alumnos universitarios, puede seleccionar su muestra acudiendo a distintas aulas de su Universidad en donde encontrará varios grupos de estudiantes. El investigador en su recorrido por distintas facultades puede ir seleccionando, para su muestra, a uno u otro grupo de alumnos a cuyas aulas llegue. Si se tiene, en un Hospital, un conjunto de Historias Clínicas, el investigador puede coger, al azar, el número de Historias Clínicas que desee o las que encuentre más a la mano.

Cuando el investigador trabaja con poblaciones pequeñas pero heterogéneas, es preferible hacer censos y evitar aplicar la técnica de muestreo. En todo caso, lo que debe hacer el investigador es ubicar las llamadas “unidades típicas” de esa población. Estas unidades típicas son las que, a juicio del investigador, más se acercan al promedio de la población. Por ejemplo, en un aula de 40 estudiantes, la muestra puede estar constituida por los alumnos que tienen rendimientos académicos medios, de ninguna manera por los que tienen rendimientos académicos muy altos o muy bajos.

En la investigación de la conducta, como quiera que se deben hacer mediciones, entrevistas, encuestas o aplicar otro tipo de instrumentos, es necesario que los sujetos que constituyan la muestra sean voluntarios, es decir, que demuestren su predisposición de participar en la investigación. En otras ciencias, como por ejemplo la Biología, esta condición no es relevante, porque si por ejemplo se estudia el comportamiento de los animales, el investigador puede coger, de las distintas jaulas de animales que dispone, el ejemplar cuyas características desea observar.

2. DETERMINACIÓN DEL TAMAÑO DE LA MUESTRA

No existe un criterio definido acerca del tamaño recomendable que debe tener la muestra. La experiencia del investigador, los recursos que dispone o las facilidades técnicas que se le presenten, serán los determinantes para fijar el tamaño de la muestra. Sin embargo, desde la etapa de planificación de la investigación, se debe fijar el tamaño de la muestra. Para ello se debe sopesar las ventajas y desventajas de trabajar con una muestra de una determinada magnitud. En principio, una muestra grande, produce información más confiable, cuando la población es relativamente pequeña. Por el contrario, una muestra pequeña extraída de una población grande, proporciona información más confiable de esta población y su grado de representatividad y generalización es mayor.

Existen muchas técnicas para calcular el tamaño de una muestra. Cada una de ellas parte de criterios específicos derivados de las características que tienen las investigaciones que se desean realizar. En términos generales, el tamaño de una muestra debe determinarse en función de las siguientes condiciones:

- La desviación estándar de los datos de la variable dependiente.
- El nivel de significación elegido para aceptar o rechazar la hipótesis nula.
- La magnitud que asigna el investigador a la diferencia de los promedios entre el grupo experimental y el de control.

Por ejemplo se trata de determinar el tamaño de la muestra necesaria para probar la siguiente hipótesis: *Existen diferencias significativas entre los promedios del grupo de profesores que participó en un proceso de capacitación profesional presencial, con respecto del grupo de profesores que participó en el mismo proceso de capacitación profesional mediante la modalidad de educación a distancia.* En este caso, luego de analizar los resultados de la prueba de rendimiento académico, se debe hallar la desviación estándar de estos datos que, por ejemplo, puede ser 3,5. Seguidamente se establece el nivel de significación en 0,05, es decir el nivel de confianza o margen de error aceptable para decidir sobre la hipótesis nula. Finalmente se establece la diferencia que se estima deben tener los promedios de los grupos, por ejemplo 2,76.

Con estos datos, es posible resolver la siguiente fórmula:

$$n = \frac{P \times 2 \times \sigma^2 \times t}{D^2}$$

En donde:

n = tamaño estimado de la muestra.

σ = desviación estándar de los datos de la variable dependiente.

t = nivel de significación para decidir acerca de la hipótesis nula.

P = tamaño de la población.
D = estimación de la diferencia entre grupos establecida por el investigador.

Al sustituir los términos en la fórmula se tiene:

$$n = \frac{300 \text{ x } 2 \text{ x } \left(3{,}5^2\right) \text{ x } 0{,}05}{2{,}76^2} = 482$$

El tamaño de la muestra, según esta técnica es: 482

Otra fórmula que permite determinar el tamaño de la muestra es la siguiente:

$$n = \frac{E \text{ x } N \text{ x } P \text{ x } Q}{E^2 \left(N - 1\right) + E \text{ x } P \text{ x } Q}$$

En donde:

n: tamaño de la muestra.
E: margen de error.
P y Q: probabilidades de éxito o fracaso (50%).
N: Tamaño de la población.
E^2 : Margen de error al cuadrado.

Reemplazando los valores, los datos son los siguientes:

$$n = \frac{4 \text{ x } 6{,}250 \text{ x } 50 \text{ x } 50}{16\left(6{,}250 - 1\right) + 4 \text{ x } 50 \text{ x } 50}$$

Hechas las operaciones resulta:

$$n = \frac{62{,}500{,}00}{109} = 568$$

El tamaño de la muestra es de 568, para una población de 6,250 sujetos.

Sin embargo, estos procedimientos complican el trabajo del investigador, por lo que resulta muy práctica la tabla de Arkin y Colton (Ver Cuadro N° 3) que permite identificar el tamaño de la muestra para una determinada población según el margen de error con el que se desea trabajar. Empleando esta tabla, se puede hallar el tamaño de la muestra deseada. Por ejemplo, con ayuda de esta tabla, si la población es 3,000, la muestra será 353, con un margen de error de 0,05, que es igual al nivel de significación elegido para decidir sobre la hipótesis nula. En el ejemplo propuesto, luego de hacer las operaciones se obtiene que, para una población de 6,250, con un margen de error de 0,04, la muestra debe ser de 568. Este dato hallado es muy cercano al valor que se encuentra en la tabla: para una población de 6,000 sujetos, con el 0,04 de margen de error, la muestra es 566.

Como se podrá observar en esta tabla, cuanto más pequeña es la población, más grande es el tamaño de la muestra. Así, para una población de 500, con un margen de error de 0.05, la muestra es del mismo tamaño, es decir del 100%. Sin embargo, para una población muy grande, de 10,000, la muestra es de 350, que viene a constituir sólo el 3,5 % de la población. Por eso no se puede afirmar que el tamaño de una muestra para que sea representativa debe ser del 5% o del 10%.

En los casos en los que el casillero aparece en blanco, el tamaño de la muestra deberá ser de más de la mitad del tamaño de la población.

TAMAÑO DE LA MUESTRA

Amplitud población	Amplitud de la muestra según márgenes de error					
	+ 1%	+2%	+ 3%	+ 4%	+ 5%	+10%
500	--	--	--	--	222	83
1,000	--	--	--	385	286	91
2,000	--	--	714	476	333	95
3,000	--	1,364	811	517	353	97
4,000	--	1,538	870	541	364	98
5,000	--	1,667	909	556	370	98
6,000	--	1,765	938	566	375	98
7,000	--	1,842	959	574	378	99
8,000	--	1,905	976	580	381	99
10,000	5,000	2,000	1,000	588	385	99
15,000	6,000	2,143	1,034	600	390	99
20,000	6,670	2,222	1,530	606	392	100
25,000	7,143	2,273	1,064	610	394	100
50,000	8,333	2,381	1,087	617	397	100
100,000	9,091	2,439	1,099	621	398	100
Infinito	10,000	2,500	11,111	625	400	100

Arkin y Colton: Tables for statisticians. Fundamental Statistics in Psychology and Education. Tokio, Mc Graw-Hill, 1965.

Cuadro Nº 3

Es muy relativo afirmar que una muestra sea grande o pequeña. Si la población es pequeña, por ejemplo 800 sujetos, la muestra podría ser del 50%, muestra porcentualmente grande, pero con sólo 400 sujetos. En cambio si la población es de 80,000, una muestra de 400 sujetos es porcentualmente muy pequeña: sólo el 0.5%. Por eso, en la práctica, los investigadores aplican la siguiente regla: muestras grandes para poblaciones pequeñas y muestras pequeñas para poblaciones grandes.

Según lo que se puede observar en la tabla y las diferentes situaciones en las que se determina el tamaño de la muestra, se establece que una población es pequeña cuando el número de sus elementos no sobrepasa los 500.

El tamaño de la muestra depende de la precisión con la que el investigador desea conocer el fenómeno que estudia. Si por ejemplo un investigador desea analizar el *nivel de rendimiento académico* de los estudiantes de una Universidad, el tamaño de la muestra dependerá de cuan exactamente este investigador desea conocer cuántos estudiantes tienen rendimiento académico alto y cuántos tienen rendimiento académico bajo en su población de estudio. En las Ciencias Sociales se acepta un margen de error de más o menos 5%. Así, si el porcentaje de estudiantes con rendimiento académico alto se calcula en el orden del 78%, luego de estudiar una muestra pequeña, el porcentaje de estudiantes con rendimiento académico alto de toda la población estará entre el 73%, que resulta de restar 5 de 78 y el 83%, que resulta de sumar 5 a 78.

Como se ve, el margen de error que se elige es arbitrario. Pero es muy importante tenerlo identificado al momento de decidir acerca del tamaño de la muestra. Los márgenes de error de diferentes magnitudes influyen en los resultados de la investigación y en las decisiones que se adopten. Si en la investigación no es aceptable correr grandes riesgos al adoptar las decisiones, los márgenes de error deben ser más pequeños, en cambio si las decisiones no conllevan riesgos serios, los márgenes de error pueden sen más grandes. Por ejemplo, si se trata de hacer presupuestos y calcular la cantidad de dinero que se va a gastar en un determinado proyecto, la decisión debe tomarse de modo muy preciso, mientras que si se va a decidir acerca de si se aplica o no se aplica un determinado método didáctico, la decisión no es tan riesgosa y el margen de error tolerable puede ser más grande.

3. ERROR DE MUESTREO

En la investigación de la conducta es más difícil que una muestra represente fielmente las características de la población. A las discrepancias entre las características de la población y las características de la muestra se denomina error de muestreo que, si es muy grande, tergiversa el sentido del concepto de muestra, pero si es pequeño, puede ser tolerado y, en tal caso, se puede asumir que una muestra representa fielmente a la población pero con un error pequeño o poco significativo que no es tan grande como para tergiversar el conocimiento de dicha población.

Por ejemplo, si el promedio del coeficiente intelectual de una población de 2,300 estudiantes universitarios es 110, una muestra será muy representativa si el promedio

del coeficiente intelectual de los 100 estudiantes de la muestra fuese también 110. En este caso no habría error de muestreo. Sin embargo, en la realidad, casi nunca se presenta esta situación y antes bien es posible encontrar coeficientes intelectuales de 109, 111 ó 110,5. Lo que interesa es establecer cuál es el margen de error que debe ser tolerado al identificar muestras para la investigación de la conducta. Baste decir que cuanto menor sea el margen de discrepancia, mayor será la capacidad de la muestra de representar fielmente a la población.

4. VENTAJAS DE LA TÉCNICA DE MUESTREO

En la investigación, la técnica de muestreo se aplica cuando es necesario y no en todos los casos. Si por ejemplo al definir la población el investigador establece que el número de los elementos de la población es 100, entonces no corresponde trabajar con una muestra, porque la población es pequeña y el investigador debe realizar un censo de todos los elementos de la población. Los datos que obtenga luego del censo se denominan, como se ha dicho, parámetros y son datos más exactos y precisos que las cifras estimadas que se obtienen por medio del muestreo.

Al realizar investigaciones no siempre se debe emplear la técnica del muestreo. Sólo debe emplearse en casos necesarios, es decir, cuando el número de los elementos de la población, definida por el investigador, sea mayor de 500 elementos. Cuando es necesario, el empleo de la técnica de muestreo proporciona considerable ayuda al investigador y le ofrece las siguientes ventajas:

Ocasiona menores costos: Los menores costos son producto del menor esfuerzo que se realiza para obtener los datos; pues el trabajo se facilita al estudiar una pequeña porción de la población. Si, por el contrario, el investigador tuviera que realizar un censo, los esfuerzos y los costos serían mayores. Paradógicamente, cuando las poblaciones son muy grandes es recomendable trabajar con muestras pequeñas.

Permite mayor rapidez: Al trabajar con pequeñas cantidades, los datos se pueden obtener más rápidamente. Si el investigador requiere la información con urgencia, el muestreo le resuelve este problema.

Proporciona mayores posibilidades de aplicación: Las técnicas censales requieren personal altamente calificado, planificación rigurosa y anticipada del trabajo, equipos especializados, etc. En cambio, las técnicas de muestreo flexibilizan el trabajo porque no requieren las condiciones exigidas cuando se trata de hacer censos, y, por tanto, proporcionan mayores posibilidades de aplicación.

Produce mayor exactitud en el estudio de la población: Cuando se reduce el número de casos de estudio, se ahorra un tiempo que el investigador puede

aprovechar para capacitar más intensamente a un menor número de personal auxiliar que realizará un mejor trabajo. Así también el investigador tendría mejores oportunidades de realizar supervisiones más minuciosas, tanto durante el trabajo de campo como en el procesamiento de la información.

5. MUESTREO PROBABILÍSTÍCO

El muestreo, según las condiciones en las que se realice, puede ser probabilístico o no probabilístico Se denominan muestras probabilísticas a las que se obtienen por procedimientos del azar o de la probabilidad, mientras que las muestras no probabilísticas se obtienen siguiendo criterios seleccionados por el investigador y por ello se llaman muestras intencionadas. Para comprender mejor esta distinción es necesario explicar qué es la probabilidad.

La probabilidad es la situación que permite seleccionar a un miembro individual de una población sin que intervengan, en absoluto, la intención o los propósitos del investigador y, por el contrario, esta selección se realice completamente al azar o en forma totalmente aleatoria. Por ejemplo, si en un programa de maestría existen 2,345 estudiantes y se desea obtener una muestra de 334 estudiantes, la probabilidad de seleccionar un estudiante para que conforme la muestra será: 334: 2,345 = 0,14. Para que un hecho sea considerado probabilístico o del azar se requiere que se cumplan las siguientes condiciones:

1º Igualdad de oportunidades y,

2º Independencia en la ocurrencia de los eventos.

La igualdad de oportunidades exige que todos los sujetos de la muestra tengan la misma oportunidad que los demás de ser seleccionados para conformar la muestra. Por ejemplo, si la población es de 30 y se quiere obtener una muestra de 1, cada sujetos tendrá una oportunidad sobre 30. En una situación auténticamente probabilística, no se acepta que uno de los 30 sujetos tenga más de una oportunidad o que alguno de los 30, no tenga ninguna oportunidad de ser seleccionado para la muestra.

La independencia en la ocurrencia de los eventos, significa que la ocurrencia de un evento no anticipa la ocurrencia del siguiente o no predisponga al investigador a seleccionar al siguiente elemento de la muestra. Si por ejemplo, en un sorteo, sale premiado el número 15, este hecho no debe anticipar que el siguiente número premiado fuese el 16. Si sucediera este evento, se puede decir que los premios están saliendo en orden correlativo, lo que descalificaría de probabilístico a dicho procedimiento muestral.

Como consecuencia de la aplicación del muestreo probabilístico se obtiene que las muestras que se extraigan tengan características muy parecidas a las de la población. Lo contrario a estos casos es la intencionalidad o el sesgo. Azar e intención

son, en consecuencia, dos extremos de un mismo "*continuum*" semántico. Azar e intención nos interesan, por igual, en la investigación de la conducta. Así, cuando los eventos suceden por obra del azar, se dice que el muestreo es probabilístico. En caso contrario, si los eventos se producen por obra de la intención del investigador, el muestreo se denomina intencionado o no probabilístico.

5.1 Tipos de muestreo probabilístico

En la investigación científica es más común y recomendable el uso del muestro probabilístico porque existe mayor posibilidad de que la muestra sea más representativa de la población. Los tipos de muestreo probabilístico más conocidos son el sorteo, el muestreo aleatorio simple, el muestreo sistemático y el muestreo estratificado.

a) Sorteo

El sorteo es el tipo de muestreo probabilístico más conocido. Para realizar un sorteo es necesario que todos los sujetos de la población estén identificados por un número y cuando se realice el sorteo, todos los sujetos deben tener la misma oportunidad de salir premiados (igualdad de oportunidades). Y si hay más de un premio, el primer sujeto premiado, no debe constituirse en indicador para intuir sobre quién recaerá el segundo premio; esto es lo que se ha denominado independencia en la ocurrencia de los eventos.

En muchos casos, los investigadores, para identificar sus muestras, hacen sorteos entre los elementos de la población, pero como es muy fácil en un sorteo sesgar los datos, la técnica del sorteo es poco recomendable.

b) Muestreo aleatorio simple

Para identificar una muestra empleando la técnica del muestreo aleatorio simple, en primer lugar es preciso conocer el tamaño de la población de la cual se desea obtener la muestra; en segundo lugar, se debe tener los listados de todos los miembros de la población; en tercer lugar, cada miembro de la población debe recibir un número de orden o un código numérico, y finalmente, se debe disponer de tablas de números aleatorios. Estas tablas son una secuencia de dígitos generados en el más completo desorden, es decir, sin que entre ellos exista ninguna correlación. En todo caso, mediante procedimientos de cómputo, se puede generar dígitos del 0 al 9, cuidando que entre ellos no exista correlación, es decir que la correlación sea cero o cercana a cero. Las tablas de números aleatorios aparecen en las páginas finales de los tratados formales de Estadística, de uno de los cuales hemos extraído parte del millón de dígitos que publicó Rand Corporation en 1955, y que aparecen en el Cuadro N° 4.

TABLA DE NÚMEROS ALEATORIOS

00000	10097	32533	76520	13586	34673	54876	80959	09117	39292	74945
00001	37542	04805	64894	74296	24805	24037	20636	10402	00822	91665
00002	08422	68953	19645	09303	23209	02560	15953	34764	35080	33606
00003	99019	02529	09376	70715	38311	31165	88676	74394	04436	27659
00004	12807	99970	80157	36147	64032	36653	98951	16877	12171	76833
00005	66065	74717	34072	76850	36697	36170	65813	39885	11199	29170
00006	31060	10805	45571	82406	35303	42614	86799	07439	23403	09732
00007	85269	77602	02051	65692	68665	74818	73053	85247	18623	88579
00008	63573	32135	05325	47048	90553	57548	28468	28709	83491	25624
00009	73796	45753	03529	64778	35808	34282	60935	20344	35273	88435
00010	98520	17767	14905	68607	22109	40558	60970	93433	50500	73998
00011	11805	05431	39808	27732	50725	68248	29405	24201	|52775	67851
00012	83452	99634	06288	98033	13746	70078	18475	40610	68711	77817
00013	88685	40200	86507	50401	36766	67951	90364	76493	89609	11062
00014	99594	67348	87517	64969	91826	08928	93785	61368	23478	34113
00015	65481	17674	17468	50950	58047	76974	73039	57186	40218	16544
00016	80124	35635	17727	08015	45318	22364	21115	78253	14385	53763
00017	74350	99817	77402	77214	43236	00210	45521	64237	96286	02655
00018	69916	26803	66252	29148	36936	87203	76621	13990	94400	56418
00019	09893	20505	14225	68514	46427	56788	96297	78822	54382	14598
00020	91499	14523	68479	27686	46162	83554	94750	89923	37089	20048
00021	80336	94598	26940	36858	70297	34135	53140	33340	42050	82341
00022	44104	81949	85157	47954	32979	26575	57600	40881	22222	06413
00023	12550	73742	11100	02040	12860	74697	96644	89439	28707	25815
00024	63606	49329	16505	34484	40219	52563	43651	77082	07207	31790
00025	61196	90446	26457	47774	51924	65394	59593	59593	42582	60527
00026	15474	45266	95270	79953	59367	83848	82396	10118	33211	59466
00027	94557	28573	67897	54387	54622	44431	91190	42592	92927	45973
00028	42481	16213	93344	08721	16868	48767	03071	12059	25701	46670
00029	23523	78327	73208	89837	68935	91416	26252	29663	05522	82562

Tomado de: Haber/Runyon: Estadística General. México, Fondo Educativo Interamericano, 1973.

Cuadro N° 4

Para identificar una muestra empleando la tabla de números aleatorios, se requiere conocer el tamaño de la población 'P' y el tamaño de la muestra 's'. Si el

tamaño de la población es una cifra de 3 dígitos, los números aleatorios se leerán de 3 en 3. Si el tamaño de la población es una cifra de 4 dígitos, los números se leerán de 4 en 4, y así respectivamente según los dígitos que tenga el tamaño de la población. La lectura de los números se puede realizar partiendo de cualquier fila o de cualquier columna de la tabla y se puede realizar la lectura comenzando, no necesariamente del primer dígito sino de cualquier dígito de la fila o columna seleccionada. Para proceder a la lectura de las tablas de números aleatorios, se utilizan varios métodos. Se explica a continuación dos de los más usados.

Método de lectura normal: Este método se emplea cuando el primer dígito del número de la población 'P' se encuentra entre 5 y 9. Así, si la población es de 857 sujetos y se requiere una muestra de 10, se empezará a leer la tabla de 3 en 3, comenzando de cualquier fila o cualquier columna de la tabla de números aleatorios hasta identificar a los 10 sujetos que se requieren para la muestra. Para ejemplificar este procedimiento presentamos una parte de la tabla de números aleatorios que aparece en el Cuadro N° 4 y si por ejemplo se empieza la lectura de la fila 18, los sujetos de la muestra serán los siguientes:

177, 435, 099, 817, 774, 027, 721, 443, 236 y 002.

Que en el fragmento de la tabla se han resaltado con negritas:

00015	65481	17614	17468	50950	58047	76974	73039	57186	40218	16544
00016	80124	35635	17727	08015	45318	22364	21115	78253	14385	53763
000**17**	**74350**	**99817**	**77402**	**77214**	**43236**	**002**10	45521	64237	96286	02655
00018	69916	26803	66252	29148	36936	87203	76621	13990	94400	56418
00019	09893	20505	14225	68514	46427	56788	96297	78822	54382	14598

Con estos mismos datos y leyendo por columnas, se procede del siguiente modo: si se comienza por las columnas 16, 17 y 18, se recorrerá, hacia abajo, las tres columnas mencionadas seleccionando los 10 primeros números distintos, comprendidos entre el 001 y 857. Los números así encontrados son los siguientes:

765, 648, 196, 093, 801, 340, 455, 020, 053 y 035.

que en el fragmento de la tabla se han resaltado en negritas:

76520	13586	34673
64894	74296	24805
19645	09303	23209
09376	70715	38311
80157	36147	64032

34072 76850 36697
45571 82406 68665
02051 65692 68665
05325 47048 90553
03529 64778 35808

14905 68607 22109
39808 27732 50725
06288 98033 13746
86507 50401 36766
87517 64969 91826

17468 50950 58047
17727 08015 45318
77402 77214 43236
66252 29148 36936
14225 68514 46427

Se sobreentiende que los números comprendidos entre el 858 y 999, no se consideran en la muestra puesto que no existen sujetos identificados con estos números dentro de este rango.

Si al terminar de recorrer la columna no se ha terminado de identificar a los 10 sujetos que se necesitan, se puede pasar a recorrer las tres columnas siguientes, o bien saltar a otras tres columnas distintas.

Método de lectura con sustracción: Es recomendable emplear este método cuando el primer dígito de la población es 4 o menor que 4. Por ejemplo, si se desea obtener una muestra de 13 sujetos de una población es 167. Esta cifra se redondea a la centena siguiente, así se sustrae 200 a cada número comprendido entre 201 y 400; se sustrae 400 a todos los números comprendidos entre 401 y 600; 600 de todos los números entre 601 y 800; y 800 de todos los números entre 801 y 999. Todos los residuos mayores de 168 hasta 200, se desechan. Por la razón anteriormente expuesta, al leer en la tabla las columnas 26, 27 y 28, se ubican los siguientes números que corresponden a los sujetos de la muestra:

346 - 200 = **146**
248 - 200 = **048**
232 - 200 = **032**
383 - 200 = **183**
640 - 600 = **040**

366 - 200 = **166**
353 - 200 = **153**
686 - 600 = **086**
905 - 800 = **105**
358 - 200 = **158**
221 - 200 = **021**
507 - 400 = **107** y
137.

Los números hallados se ubican en el siguiente fragmento de la tabla:

00000	10097	32533	76520	13586	**346**73	54876	80959	09117	39292	74945
00001	37542	04805	64894	74296	**248**05	24037	20636	10402	00822	91665
00002	08422	68953	19645	09303	**232**09	02560	15953	34764	35080	33606
00003	99019	02529	09376	70715	**383**11	31165	88676	74394	04436	27659
00004	12807	99970	80157	36147	**640**32	36653	98951	16877	12171	76833
00005	66065	74717	34072	76850	**366**97	36170	65813	39885	11199	29170
00006	31060	10805	45571	82406	**353**03	42614	86799	07439	23403	09732
00007	85269	77602	02051	65692	**686**65	74818	73053	85247	18623	88579
00008	63573	32135	05325	47048	**905**53	57548	28468	28709	83491	25624
00009	73796	45753	03529	64778	**358**08	34282	60935	20344	35273	88435
00010	98520	17767	14905	68607	**221**09	40558	60970	93433	50500	73998
00011	11805	05431	39808	27732	**507**25	68248	29405	24201	52775	67851
00012	83452	99634	06288	98033	**137**46	70078	18475	40610	68711	77817
00013	88685	40200	86507	50401	36766	67951	90364	76493	89609	11062
00014	99594	67348	87517	64969	91826	08928	93785	61368	23478	34113

Con la misma población y muestra, si se comienza a leer la tabla a partir de la vigésima fila, los números son los siguientes:

190
989 - 800 = **189**
320 - 200 = **120**
505 - 400 = **105**
142
256 - 200 = **056**
851 - 800 = **051**
464 - 400 = **064**
275 - 200 = **075**

678 - 600 = **078**
896 - 800 = **096**
297 – 200 = **097**
788 – 700 = **088**

Que en el siguiente fragmento de la tabla se encuentran en la siguiente ubicación:

00018	69916	26803	66252	29148	36936	87203	76621	13990	94400	56418
000**19**	**09893**	**20505**	**14225**	**6851**4	**46427**	**56788**	**96297**	**788**22	54382	14598
00020	91499	14523	68479	27686	46162	83554	94750	89923	37089	20048
00021	80336	94598	26940	36858	70297	34135	53140	33340	42050	82341
00022	44104	81949	85157	47954	32979	26575	57600	40881	22222	06413

Si la población fuese 384, se redondea la cifra a 400 y se sustrae 400 a todo número comprendido entre 401 y 800 y automáticamente se desechan todos los números mayores de 800.

Si se requiere extraer varias muestras de una misma población, es aconsejable cambiar el punto de partida en la tabla. Así será posible disponer de varias series de muestras y, eventualmente, se pueden identificar muestras de muestras.

c) Muestreo sistemático

El muestreo sistemático es una técnica diferente de la anterior. Para utilizar esta técnica se requiere disponer de listados de los sujetos que deben estar identificados por sus nombres o por sus códigos numéricos. Es preciso conocer el tamaño de la población 'P' y el tamaño de la muestra 's'. Sea, por ejemplo 750, el tamaño de la población y 15, el tamaño de la muestra. Entonces se divide P entre s. El coeficiente hallado es la constante 'k'. En este caso:

$$K = P / s$$

ó sea:

$$750 / 15 = 50.$$

La constante así encontrada es 50. Ubicada la constante, se recorre los listados en los que aparecen los nombres de los sujetos de la población y se entresaca, para la muestra, a los sujetos que se hallan separados por cada 50 nombres, es decir, por la distancia establecida por la constante. Pero en este caso, si se comienza a identificar la muestra a partir del sujeto que tiene el número 01, y si se sabe cuál es la constante, va a ser posible anticipar que los sujetos de la muestra serán quienes cuyos números de orden sean múltiplos de 50 (50, 100, 150, 200, etc.) Este hecho

estaría en contradicción con el segundo principio que rige los fenómenos probabilísticos: la independencia en la ocurrencia de los eventos, porque la ocurrencia del primer evento, ya está anticipando la ocurrencia de los siguientes. Para evitar esta situación, se sortea los números ubicados en el primer rango, en este caso, los primeros 50, (la constante) y se comienza a aplicar la constante a partir del número que salió sorteado. Por ejemplo, si en el sorteo de los 50 primeros números sale el número 13, se aplicará la constante a partir del 13 y se procederá de la siguiente manera:

$$13 + 50 = 63,\ 63 + 50 = 113,\ 113 + 50 = 163\ ;$$

y así sucesivamente hasta agotar la lista de los sujetos de la población.

Este es un método que tiene la gran ventaja de entresacar a los sujetos de la muestra de manera uniforme en toda la población, es decir, este método permite distribuir la muestra de manera uniforme en toda la población.

Otra ventaja de este método es que se puede identificar a los sujetos de la muestra sin cometer errores. Además es más rápido en algunos casos como por ejemplo si se tienen datos en un fichero, y la constante es una determinada cifra y si el grosor de las fichas que contiene la constante es, por ejemplo, de una pulgada, se puede extraer una ficha, sin necesidad de contar cada una de las fichas, sino simplemente midiendo, de pulgada en pulgada, el espacio que ocupan las fichas. Naturalmente que con esta variante del método no se identificará exactamente a los sujetos de la muestra, pero la aproximación será muy alta.

Otra situación puede ser la siguiente. Si por ejemplo se desea obtener una muestra de 400 personas cuyos nombres aparecen en la Guía Telefónica de Lima, se puede proceder de la siguiente manera: Si la Guía Telefónica de Lima tiene 1,000 páginas y en cada página aparecen cuatro columnas de nombres y en cada columna hay 135 nombres, entonces su puede calcular, rápidamente el tamaño de la población, en este caso el número de abonados al servicio telefónico en Lima. El resultado de estos cálculos es el siguiente:

$$135 \times 4 \times 1{,}000 = 540{,}000$$

La población es, entonces, 540,000 abonados. Si se desea obtener una muestra de 400, se dividirá 540,000 entre 400, lo que dará la constante, K, o sea 1,350. Como quiera que en cada página aparecen 540 nombres (135 por columna por cuatro columnas de nombres), se dividirá 1,350 entre 540, lo que da la cifra de 2,5, es decir, se sacará un nombre cada 2,5, dos páginas y media de la Guía Telefónica y así, al terminar de hojear dicha Guía, se habrá identificado a las 400 personas que se requieren para la muestra. Si se desea todavía mayor precisión, se comenzará identificando el primer nombre luego de hacer un sorteo de las 2,5 primeras páginas

de la Guía. Pero si se va a sacar un nombre cada 2,5 páginas de la Guía, ¿de cuál de estas 2,5 páginas se sacará el nombre? Lo aconsejable en este caso es sacar el nombre del centro, es decir de los lugares centrales de cada 2,5 páginas.

Lo dicho tiene sus consecuencias en la precisión al identificar el tamaño de la muestra. Al parecer, este es un método más exacto, pues entresacando a un sujeto luego de cada 2,5 páginas, es decir según la constante, se puede decir que la muestra se ha distribuido uniformemente en toda la población.

d) Muestreo estratificado

Si al investigador sólo le interesan los sujetos en términos de cantidad para identificar las muestras es suficiente emplear las técnicas que acabamos de describir. En cambio, si el investigador está interesado en conocer algunas características de la población, tales como edad, sexo, etc., es necesario estratificar la población y lograr que las características de la muestra coincidan con las características estratificadas de la población. El muestreo estratificado consiste en reducir la población en función de los sub conjuntos o estratos que se han identificado en ésta. Sea el caso de la población de estudiantes universitarios de la ciudad de Lima y se desea estratificar la población según dos variables: sexo de los estudiantes y tipo de gestión de la Universidad. Así se tendrá, si cada variable asume dos valores, los cuatro siguientes estratos: 1. Estudiantes varones de universidades nacionales; 2. Estudiantes varones de universidades particulares; 3. Estudiantes mujeres de universidades nacionales y 4. Estudiantes mujeres de universidades particulares. Estos cuatro estratos, tienen sus respectivos tamaños, no son todos iguales, y el investigador debe conocer el tamaño de cada estrato.

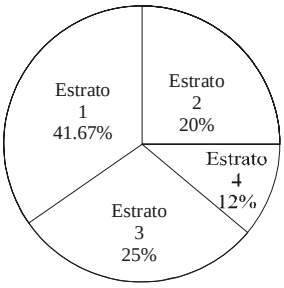


Gráfico 4

Si estos son los estratos y se conoce el tamaño de cada uno, el investigador tendrá que representar, en la muestra y en los mismos porcentajes, cada uno de los estratos de la población. Luego de establecidos los estratos en la muestra para

identificar a los sujetos se puede emplear cualquiera de los procedimientos descritos anteriormente. Lo que en realidad se hace es reducir el tamaño de la población, guardando su apariencia estratificada original.

Sea el siguiente ejemplo:

Total de estudiantes varones y mujeres de universidades nacionales y particulares:	120,000	100,00 %
Total de estudiantes de universidades nacionales:	80,000	66,67 %
Total de estudiantes de universidades particulares:	40,000	33,33 %
Total de estudiantes varones de universidades nacionales:	50,000	41,67 %
Total de estudiantes mujeres de universidades nacionales:	30,000	25,00 %
Total de estudiantes varones de universidades particulares:	25,000	20,83 %
Total de estudiantes mujeres de universidades particulares:	15,000	12,50 %

Con estos datos se puede hacer la siguiente graficación:

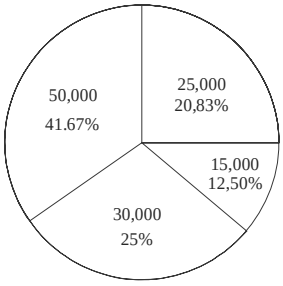


Gráfico N° 5

Como se trata de obtener una muestra de 400 sujetos, entonces se procederá, respetando los porcentajes encontrados, a calcular el número de sujetos para cada estrato. Los datos son los siguientes:

Muestra de estudiantes varones de universidades nacionales:	41,70 % =	167
Muestra de estudiantes mujeres de universidades nacionales:	25,00 % =	100
Muestra de estudiantes varones de universidades particulares:	20,83 % =	83
Muestra de estudiantes mujeres de universidades particulares:	12,50 % =	50

Esto significa que de los 50,000 estudiantes varones de universidades nacionales, se debe elegir a 167 estudiantes para la muestra. De las 30,000 estudiantes mujeres de universidades nacionales, se debe elegir a 100 estudiantes para la muestra. De los 25,000 estudiantes varones de universidades particulares, se debe elegir a 83

estudiantes para la muestra y de las 15,000 estudiantes mujeres de universidades particulares, se debe elegir a 50 estudiantes para la muestra.

Pero ¿cómo identificar a estos estudiantes en la población? Para estos casos se aplican las técnicas probabilísticas descritas. Por ejemplo, mediante el muestreo aleatorio simple se puede identificar la muestra de 167 sujetos de una población de 50,000. Así se debe proceder en cada uno de los estratos. Esta técnica se denomina muestreo aleatorio estratificado.

Las ventajas de aplicar la técnica del muestreo aleatorio estratificado son las siguientes:

Si se requiere precisión en los datos que se analizan en cada una de las divisiones de los estratos, se puede considerar una población independiente a cada uno de los estratos. Por ejemplo, el número de estudiantes de universidades nacionales.

También es conveniente la estratificación por razones administrativas: Si una institución tiene, por separado, identificados a los alumnos varones y mujeres, los estratos identificados facilitan el manejo de los datos institucionales.

Los sujetos de la muestra pueden ser diferentes en función de cada estrato. El investigador puede asumir que las características de los estudiantes de universidades nacionales no son las mismas que las características de los estudiantes de universidades particulares. Para mantener sus respectivas identidades o particularidades, es conveniente estudiarlas en estratos diferentes.

Al identificar estratos, el investigador puede obtener mayor precisión en la estimación de las características de cada estrato, porque si la población es heterogénea, los estratos ya son homogéneos, pues se han agrupado a los miembros de cada estrato por sus características comunes: estudiantes varones, que obviamente tienen características diferentes a las que poseen las estudiantes mujeres.

El investigador, en función de las necesidades que tenga de analizar los datos o, en función de la precisión que busca, debe estratificar la población, según los criterios que crea más convenientes. Asimismo, depende de cada investigador determinar el tamaño de la muestra en función del tamaño del estrato que tiene identificado.

Por lo general, la varianza de las muestras estratificadas es más pequeña que la varianza que se observa en las muestras obtenidas con la técnica del muestreo aleatorio simple. Esta menor varianza ya constituye una importante ventaja de este método.

6. MUESTREO NO PROBABILÍSTICO

Cuando el investigador, por las características de su estudio, necesita incorporar en la muestra a sujetos que posean las características deseadas por él, no es posible emplear el muestreo probabilístico, sino un tipo de muestreo completamente distinto. Este es un tipo de muestreo en el que la probabilidad no es imprescindible y, por el contrario, es necesario identificar a los sujetos de la muestra con criterios intencionados, es decir, identificando a los sujetos de la muestra en base a criterios previamente establecidos. Si por ejemplo se trata de un concurso universitario de ajedrez, y las autoridades de una determinada Universidad deben acreditar a tres representantes, entonces seleccionarán a los tres mejores jugadores de ajedrez que tiene la Universidad para que la representen en el concurso. Como es natural, a estos tres jugadores que constituyen la muestra de la Universidad, no se los puede identificar mediante el azar ni empleando los métodos probabilísticos, sino de forma intencionada buscando, entre todos los alumnos de la Universidad, a los tres estudiantes que reúnen las características necesarias, en este caso, ser buenos jugadores de ajedrez.

a) Muestreo por cuotas

Cuando se desea obtener la muestra de una población organizada en forma piramidal o arborescente, como es el caso del sistema educativo nacional, se puede asignar cuotas a cada escalón de la organización. Por ejemplo, si se desea hacer una investigación, a nivel nacional, sobre el rendimiento de los estudiantes de educación secundaria y se estima que la población estudiantil de secundaria en el Perú sobrepasa los dos millones de alumnos y se desea obtener una muestra de 504 estudiantes, se procede de la siguiente manera: Si por ejemplo, desde el punto de vista de la administración, el Ministerio de Educación tiene 8 Regiones, se asignan cuotas a cada una de las Regiones. Las cuotas pueden ser proporcionales a la población estudiantil de cada Región o bien pueden ser iguales para todas. En el caso que se decidiera que fueran iguales, se asignarían cuotas de 63 alumnos por cada Región (504 / 8 = 63). Seguidamente, en cada una de las Regiones, se asignarán nuevas cuotas para ser cubiertas por cada una de las UGEs. Suponiendo que en la Región Nº 1, existieran 7 UGEs, a cada UGE le correspondería una cuota de 9 sujetos (63 / 7 = 9). Si cada UGE tuviera 9 Colegios, se asignará un alumno por Colegio. En este caso, para identificar al alumno de cada Colegio, se tendría que hacer un sorteo entre todos los alumnos o aplicar los otros tipos de muestreo como el aleatorio simple o el muestreo estratificado. Luego, las autoridades harán el reporte inverso, de abajo hacia arriba, es decir el informe de los colegios a las UGEs, el informe de las UGEs a la Direcciones Regionales y éstas a la Sede Central de Ministerio de Educación y así se lograría, sumando las cuotas de las Regiones, obtener la muestra de los 504 estudiantes requeridos para el estudio.

b) Juicio de expertos

El juicio de expertos es el criterio que maneja un profesional especializado cuando tiene que seleccionar sujetos. El experto decide qué sujetos van a participar en la muestra. Si por ejemplo se convoca a nivel de Lima a un Concurso Universitario de Matemática, serán los matemáticos o los profesores de matemática quienes deberán emitir opinión o seleccionar, ellos, a los mejores alumnos de Matemática de cada Universidad para que representen, exitosamente, a la institución. Si se trata de estudiar aspectos relacionados con el retardo mental, los profesores de la especialidad, o mejor un psicólogo, y más aún, un equipo multidisciplinario de profesionales, actúan como expertos para diagnosticar o decidir sobre el tipo de tratamiento que requieren estos alumnos.

En estos casos, aplicando el juicio de expertos se determina quiénes van a formar parte la muestra. Las selecciones deportivas de una país, por ejemplo las de fútbol o voleybol, son muestras representativas del país y los jugadores han sido escogidos por el seleccionador/entrenador de entre un universo muy vasto de deportistas. En este caso, el seleccionador/entrenador actúa como experto para emitir opinión y convocar a los jugadores.

c) Muestreo ocasional

En muchas oportunidades, en la investigación de la conducta se recoge la opinión de las personas con las que ocasionalmente tropieza el investigador. Esta es la técnica más sencilla y muchas veces la única manera de recopilar información. Su validez es relativa aunque parte del supuesto que los sujetos a quienes se entrevista son representativos de la población. Por ejemplo, las agencias dedicadas al estudio de la opinión colectiva, para anticipar los resultados de una votación de los ciudadanos, recogen información de los electores que salen del local en el que votaron. A partir de estos informes, realizan sus progresiones para llegar a la predicción deseada.

LECTURA COMPLEMENTARIA

CONDICIONES Y ELEMENTOS DE LAS MUESTRAS

Restituto Sierra Bravo*

NOCIÓN

Una vez definido el problema a investigar, formuladas las hipótesis, determinadas las dimensiones e indicadores de las variables y delimitado el campo de la investigación, se hace preciso realizar, además de las anteriores, otra opción antes de empezar con el trabajo de campo, la recogida de datos mediante la aplicación de los métodos elegidos.

Esta nueva opción consiste en la decisión de si las unidades objeto de observación o estudio van a ser todas las que forman el universo, o únicamente se va a extender la indagación a una parte representativa o muestra de aquéllas.

Salvo en el caso de poblaciones pequeñas, esta última solución es la que se impone en la realidad, dado que en los demás casos, razones de tiempo, coste y complejidad de las operaciones de recogida, clasificación y análisis de los datos, descartan en la práctica la posibilidad de que el estudio abarque individualmente a todas las unidades que comprende el universo.

Por ello, al constituir el uso de las muestras una técnica de aplicación casi general en las investigaciones sociales, es evidente la importancia que tiene su estudio y la necesidad en que se halla el investigador social de conocer por lo menos sus principios y prácticas básicos, aunque se trate de una materia, basada en las leyes de azar y el cálculo de probabilidades, que pertenece al campo matemático de la estadística.

Una muestra es *simplemente*, en general, una parte representativa de un conjunto, población o universo, cuyas características debe reproducir en pequeño lo más exactamente posible.

De modo más científico, se pueden definir las muestras como una parte de un conjunto o población debidamente elegida, que se somete a observación científica

---

* Tomado de: Sierra Bravo, R: Técnicas de investigación social. España. Paraninfo, S.A. 1992. pp. 174-190.

en representación del conjunto, con el propósito de obtener resultados válidos, también para el universo total investigado.

Las muestras tienen un fundamento matemático estadístico. Éste consiste en que obtenidos de una muestra, elegida correctamente y en proporción adecuada, unos determinados resultados, se puede hacer la inferencia o generalización, fundada matemáticamente, de que dichos resultados son válidos para el universo del que se ha extraído la muestra, dentro de unos límites de error y probabilidad, que se pueden determinar estadísticamente en cada caso.

Las muestras presentan las siguientes ventajas, en las ciencias sociales:

1. Mediante ellas, con una muestra relativamente reducida en relación al universo, se pueden encuestar las grandes poblaciones y núcleos humanos, que de otra manera sería muy difícil o prácticamente imposible investigar.
2. En todo caso las muestras suponen una gran economía en las encuestas y la posibilidad de mayor rapidez en su ejecución.
3. Además, como señala Sheuch, en el "Tratado de Sociología empírica", de R. König (Tecnos), las muestras ofrecen otra ventaja, muy importante y casi inadvertida; una muestra puede ofrecer resultados más precisos que una encuesta total, aunque esté afectada del error que resulta de limitar el todo a una parte.

En una muestra se puede prestar más atención al caso particular. De ello resulta eventualmente una limitación de los errores fácticos de tal orden, que la muestra puede conducir a un resultado más preciso en su totalidad de una encuesta total. De ahí que se utilicen muestras en las estadísticas americanas para comprobar el grado de exactitud de las encuestas totales.

CONDICIONES DE LAS MUESTRAS

Las condiciones fundamentales de las muestras son cuatro:

1. Que comprendan parte del universo y no la totalidad de este.
2. Que su amplitud sea estadísticamente proporcionada a la magnitud el universo. Esta condición se halla en relación con el punto práctico determinación del tamaño de la muestra y sirve para decidir si, según las unidades que comprende respecto al universo, una muestra es o no admisible.
3. La ausencia de distorsión en la elección de los elementos de la muestra. Si esta elección presenta alguna anomalía, la muestra resultará por este mismo hecho viciada.
4. Que sea representativa o reflejo fiel del universo, de tal modo que reproduzca sus características básicas en orden a la investigación. Esto quiere decir que si hay sectores diferenciados en la población que se supone ofrecen características

especiales, a efectos de los objetivos de la investigación, la muestra también deberá comprenderlos y precisamente en la misma proporción, es decir, deberá estar estratificada como el universo.

Además de estas condiciones fundamentales, se ha de tener en cuenta que las muestras se basan, como han destacado algunos autores, en una teoría del público como masa atomizada (Bordieu), producen la ruptura de las redes topológicas y arrancan al individuo de la urdimbre social (Ibáñez) y prescinden de rasgos sociales estructurales importantes (Scheuch). Por ello, el empleo de las nuestras debería implicar una información previa sobre la estructura de la colectividad investigada.

Una cierta solución a este problema son los llamados sondeos contextuales como, por ejemplo, el muestreo gradual, para conocer la opinión de los militantes de un partido de, primero, federaciones provinciales, segundo, de secciones locales y, tercero, de militantes de éstas. Se ha de recordar que el individuo no está nunca totalmente desconectado de la sociedad, sino que, aunque actúe aisladamente de ésta, lleva en sí el influjo de los grupos sociales a los que pertenece.

BASE DE LA MUESTRA

Los elementos principales de la muestra son la base y la unidad de la misma.

La base de la muestra es, en todo caso, la población de la que se obtiene o saca la muestra. Esta población puede no estar censada o puede estarlo. En el primer caso, existe la base de la muestra en sentido real o material, pero no en sentido formal, mientras que en el segundo caso hay base de la muestra tanto material, la población real a investigar, como formal, el censo de todos los individuos que la componen. Precisamente con el término de base de la muestra se suele designar de manera específica este censo o registro.

La importancia de la base de la muestra, tomada específicamente o en sentido estricto, se deriva de que la muestra, operativamente, es el resultado de la elección de unidades dentro de una población o conjunto previamente determinado de aquéllas. Por ello, fundamento básico de la muestra, es la existencia de un registro de dicho conjunto, en el que aparezcan individualizadas todas sus unidades, y permita realizar la elección mediante un sorteo riguroso.

Esta concreción individualizada de las unidades del universo es el punto de partida y el fundamento necesario para realizar con rigor la elección en que consiste la muestra y por ello se dice que constituye la base de la muestra en sentido estricto.

Ésta puede consistir en un censo, un registro, una lista, un fichero, un catálogo, un mapa, un plano, etc. Se debe dar normalmente un número a cada ficha o elemento del catálogo.

En la base de la muestra deben figurar individualizadas todas las unidades que forman la población con expresión de su número en el universo, nombre, domicilio en su caso, etc. Por ello, la base de la muestra hace posible la identificación de los elementos que se hayan seleccionado mediante la muestra y su encuesta posterior.

La base de la muestra en este sentido estricto, no siempre existe. Hay muchos universos que no están censados o catalogados y que es prácticamente imposible catalogar. Por ejemplo, no lo están el público que circula por las calles ni los asistentes a un espectáculo. La solución que se adopta entonces es practicar la elección de la muestra por algún procedimiento aleatorio imperfecto, como v.g. encuestar uno de cada cinco que se encuentren por la calle

Lo ideal sería formar una base de muestra específica para la encuesta. Pero esto no es posible, por lo general, en las investigaciones sociales. Supondría casi tanta dificultad como efectuar la encuesta a toda la población. Por ejemplo, si para hacer una encuesta que alcanzase a toda España se decidiese formar una base de la muestra propia, esto supondría realizar un censo de todos los españoles.

Por esta razón, normalmente se utilizan bases ya formadas. En este caso es preciso asegurarse de que la base está actualizada, pues si no nos exponemos a dejar fuera de la muestra, por ejemplo, a la población más joven de ella.

Además hay que verificar si la población que queremos encuestar (universo teórico) y aquella de la que se toma de hecho la muestra (universo de trabajo) o a la que se refiere la base de la muestra, coinciden, porque podría haber discrepancias más o menos extensas entre ellas.

En el orden práctico, respecto a la base de la muestra, hay que buscarla, analizarla y prepararla a efectos del muestreo si existe. Si no existe, y ello es posible, hay que formarla. En este caso se debe cuidar, de modo especial, según lo indicado, que no resulte ordenada de alguna forma que pueda determinar una elección distorsionada o no representativa de la muestra.

Por último hay que advertir que algunos autores, entre ellos Sánchez Crespo (1976, p. 6), llaman a la base de la muestra marco.

Además de este sentido estricto que he indicado, entiende también por marco, en sentido amplio, la información previa, mapas, comunicaciones, etc, que permita llegar a todos los elementos de la población a encuestar.

UNIDAD DE LA MUESTRA

La unidad de la muestra es cada uno de los elementos que comprende su base y figuran, numerados e individualizados, en el registro de la misma cuando existe.

Las unidades de este registro representan y remiten a las unidades de observación que forman el universo a investigar.

Como he indicado la base formal, estricta, y sus unidades hacen posible la identificación de las unidades de observación del universo a encuestar.

Las unidades de la muestra no sólo pueden ser simples, sino también colectivas, como cuando están constituidas por familias, grupos, sectores de ciudades, pueblos, etc. El número de individuos que comprende la unidad de la muestra cuando es colectiva se llama talla de la muestra.

Esta talla de las unidades de la base de la muestra o población es lo que determina lo que se suele llamar el nivel de agregación en la investigación. Este nivel puede presentar grados de agregación diversos, bien de carácter espacial, social, económico, según los colectivos que se elijan como agregados.

Así, por ejemplo, para estudiar las compras de TV en color se puede operar al nivel de agregación de individuos o de colectivos y, dentro de los colectivos, especialmente, éstos pueden ser municipios, distritos, regiones, naciones. Desde un punto de vista social los agregados pueden ser familias, establecimientos de venta, empresas, organismos y asociaciones determinados.

RELACIÓN ENTRE UNIVERSO Y MUESTRA

El diagrama que sigue de Mora y Araujo (1973, p. 12), muestra gráficamente la relación entre universo y muestra.

La muestra, como se ve, es una parte reducida de una determinada población o universo, en relación a la cual representa cuantitativamente una fracción.

Pero también la muestra tiene la misma forma circular que el universo. Ello pretende indicar que, cualitativamente, la muestra debe tener las mismas características que la población que representa.

El universo está formado por toda la población o conjunto de unidades que se quiere estudiar y que podrían ser observadas individualmente en el estudio.

Sin embargo, por razones de economía se investiga, en su representación, la muestra.

En el gráfico también se ve cómo el universo o población forma parte también de un conjunto más amplio al que se llama universo hipotético. Este universo hipotético está constituido por todas las poblaciones existentes en las que se podría también efectuar la investigación proyectada.

Figura 1

Precisamente lo que da validez máxima a un hallazgo científico es haber sido comprobado no sólo en una población sino en relación a muchas de las existentes.

Entre el universo y la muestra se puede distinguir una relación cualitativa y otra cuantitativa.

A la relación cuantitativa se refieren los conceptos de fracción de muestreo y de coeficiente de elevación. La fracción de muestreo es el porcentaje que representa la muestra respecto al universo.

Se obtiene dividiendo la muestra por el universo. Por ejemplo, si en un universo de 3,000 unidades hemos elegido una muestra de 600, la fracción de muestreo será 0,20 o el 20%.

Esta fracción puede tener diversas aplicaciones, entre ellas la de hallar la composición por estratos de una muestra cuando se conoce la composición por estratos del universo o población.

En cuanto al coeficiente de elevación no es otra cosa que el número inverso a la fracción de muestreo y consiste en la cantidad por la que hay que elevar la muestra para obtener el universo.

Se obtiene, al revés de la anterior, dividiendo el universo por la muestra. En el caso indicado sería 5. Prácticamente el coeficiente de elevación señala las veces que la muestra está contenida en el universo.

La relación cualitativa exige, como se ha indicado, que el universo y la muestra reúnan las mismas características básicas en igual proporción. Ello es importante sobre todo en los universos estratificados, que comprenden categorías diversas.

Esta correspondencia es difícil que en la práctica se de exactamente, sobre todo porque en las encuestas siempre hay fallos: entrevistas previstas que no se pueden realizar o no son válidas, cuestionarios que no se devuelven, etc.

Por ello, para que se conozca la discrepancia existente, es importante formar un cuadro comparativo en el que consten los porcentajes que suponе cada categoría en el universo y en la muestra. A esta operación se le llama verificar el equilibrio de la muestra.

Esta operación hace ver a veces que los porcentajes de los estratos en las muestras, sobre todo una vez realizada la encuesta, difieren bastante de los del universo.

Hay casos en los que esta discrepancia se procura deliberadamente, a fin de que un estrato que interesa estudiar especialmente y se considera insuficientemente representado en la muestra lo esté más. Para ello se aumenta el porcentaje que le corresponde y se disminuye el de otros estratos en la misma proporción. En estos casos, se debe corregir esta discrepancia en los resultados, ponderando estos en la muestra según los porcentajes que corresponden a cada estrato en el universo.

Ejercicios

1. Dadas las investigaciones que se indican a continuación, se pide determinar razonadamente si se basan en una muestra del universo correspondiente y si, en caso afirmativo, la muestra es correcta.

A) A efectos de realizar el sociograma de una clase, se ha pedido a todos los alumnos que expresaran en una papeleta los nombres de sus compañeros de clase que les gustaría tener sentados a su lado y aquellos que no.

Respuesta

Es claro que no constituye una muestra, en cuanto el universo son todos los alumnos de la clase y todos ellos fueron encuestados.

B) Para estudiar las prácticas sexuales de los varones en una prisión se entrevistó a todos los que se presentaron voluntariamente a responder al cuestionario que se había preparado.

Respuesta

Genéricamente se puede decir que, en este caso, hay una muestra, en cuanto solo se hizo la encuesta a una parte del total de internados en la prisión. Sin embargo, se trata de una muestra viciada, basada en un sistema de elección inadecuado, por lo que no se puede considerar representativa del universo, ni sus resultados extendidos a este.

C) Un antropólogo social ha convivido durante dos años con una familia típica de una localidad de las Hurdes, se ha ganado su confianza y ha logrado que sus miembros le expusieran los aspectos de su vida de interés para su investigación.

Respuesta

Este estudio, aunque se diga que se basa en una familia típica, no se puede considerar como muestra, pues un solo caso no es suficiente al efecto.

D) Para estudiar las infracciones de circulación cometidas por no detención ante un "stop", un equipo de observación ha permanecido, de ocho de la mañana a ocho de la tarde, ante el "stop" durante tres días de la semana consecutivos, y los observadores consignaron los detalles de una por cada tres infracciones observadas.

Respuesta

El universo son todas las infracciones. Como sólo se investiga una parte, se puede hablar de muestra de todas ellas. Pero esta muestra es desviada y no representativa del universo, porque sólo proporciona información de unas horas determinadas y de sólo tres días consecutivos a la semana, sin que nada diga sobre las restantes horas del día y los demás días de la semana, en los que el número y calidad de las infracciones puede ser diferente.

E) Para predecir los resultados en una comunidad de una elección municipal que iba a tener lugar unos días después, el encuestador preguntó su candidato preferido a todos los hombres y mujeres con derecho a voto.

Respuesta

No existe muestra, pues se consulta a todo el universo, los hombres y mujeres con derecho a voto y no a parte de ellos. Pero, por otra parte, el universo no es representativo a los efectos de la encuesta, ya que se dice, se consulta indiscriminadamente a todas las personas con derecho a voto y hay que tener en cuenta que siempre suele haber un porcentaje a veces muy elevado que no votan. En consecuencia, sus preferencias viciarían la predicción, pues los resultados efectivos se basan únicamente en la conducta de los que realmente tengan intención de votar y voten.

F) En otra elección a diputados, en la que se presentaban los candidatos, uno de ellos de izquierdas, se realizó el sondeo de opinión, a una muestra elegida al azar con base en la lista telefónica, por medio del teléfono. Se obtuvo un resultado favorable al candidato de derechas, aunque resultó luego elegido el de izquierdas.

Respuesta

En este sondeo la muestra tampoco es representativa, aunque se haya elegido al azar. Presenta la distorsión que supone el hecho de que los que poseen teléfono son gente de un cierto nivel económico, y por ello es de suponer que sean en mayor proporción de derechas. Ello explica la falta de acierto de la encuesta del resultado efectivo de las elecciones.

G) Para descubrir la existencia de "cliques" en una clase numerosa de la universidad, el investigador preguntó a una muestra elegida al azar de estudiantes el nombre de sus amigos más íntimos.

Respuesta

Si se basa en una muestra, según el enunciado, pero esta muestra es improcedente en relación al fin propuesto en la investigación, descubrir los cliques, porque para ello es preciso que la encuesta abarque a todos los miembros de la clase.

"Clique", como es sabido, es el subgrupo cerrado de amigos íntimos que se eligen todos ellos unos a otros, recíprocamente.

H) Para estudiar las actitudes políticas de una sociedad cultural que agrupa 5,000 miembros se ha decidido realizar una encuesta por cuestionario a 500 de ellos elegidos arbitrariamente, además de 200 entrevistas a otros tantos socios elegidos al azar, si bien en la realidad los entrevistadores se permitieron sustituir frecuentemente los socios elegidos por otros.

Respuesta

Hay en los dos casos, en principio, muestra del universo. Sin embargo, la primera es inadmisible en cuanto no reúne la condición de basarse la elección en un procedimiento racional, si es posible al azar, y además se opone al principio de que dicha elección no debe ser arbitraria, pues hay gran peligro de que prevalezcan criterios subjetivos en ella. En el segundo caso la muestra correcta inicialmente ha resultado viciada en la realidad por la sustitución personal y, por tanto, subjetiva que los entrevistadores se han permitido.

I) Una empresa de mercados para investigar el interés del público mayor de edad de una ciudad donde viven 8,000 blancos, 5,000 negros y 3,000 de otras razas, por una marca de licor determinada, ha decidido realizar una encuesta a una muestra elegida al azar de 400 negros.

Respuesta

En este caso también se trata de una muestra invalida, porque aunque el universo es toda la población, sólo se basa en un sector de ella y además se da la

circunstancia que es homogéneo, a diferencia de la población, que es heterogénea. Para ser correcta la nuestra debería comprender una parte de las tres razas en la misma proporción. Es decir, debería estar estratificada.

j) Para estimar la proporción de amas de casa que habían visto un programa de TV en la semana anterior, los investigadores eligieron una muestra al azar de manzanas de la ciudad. En una de dichas manzanas, compuesta de 150 pisos, los encuestadores entrevistaron a 100 amas de casa, que fueron encontradas en el piso al llamar a la puerta.

Respuesta

Aquí se tiene que un porcentaje muy elevado de la muestra, 50 de 150, no ha podido ser entrevistado. Ello quita representatividad a la muestra, sobre todo porque muchas de las que faltaban es posible que, por trabajo u otras ocupaciones, estén habitualmente fuera de casa y, por tanto, en ellas la proporción de las que no han visto el programa sea mayor.

2. Supuestos los siguientes estudios que se ha pensado realizar por encuesta muestral, se pide indicar la base y la unidad de la muestra en todos ellos.

A) Análisis de la evolución en sus diversos aspectos de las informaciones religiosas en los periódicos españoles antes y después del Concilio.

Respuesta

En primer lugar hay que advertir que el estudio abarca una pluralidad de periódicos, de años y de números de cada periódico dentro de cada año. La muestra general debe comprender, pues, tres submuestras correspondientes a las diversas pluralidades indicadas.

Respecto a los periódicos, la base de la muestra debería ser el registro oficial de todos los periódicos españoles autorizados y la unidad de la muestra cada periódico.

En cuanto a los años, es obvio que la base de la muestra es la lista natural de todos los que abarque la investigación y la unidad cada año.

Asimismo, dentro de cada periódico y año, la base sería el total de números publicados por cada periódico al año y la unidad cada número.

B) Un estudio sobre las condiciones de viviendas familiares de la zona madrileña de Vallecas.

Respuesta

En este caso la base de la muestra debería ser el plano de la zona, en el que, numeradas correlativamente por sectores a efectos de muestreo estratificado en su caso, las manzanas constituirían la unidad de la muestra.

C) Un estudio sobre las condiciones estructurales y funcionales de las asociaciones voluntarias privadas de España, con exclusión de las económicas, religiosas, políticas y sindicales.

Respuesta

Base de la muestra: El registro oficial de asociaciones. Unidad: cada asociación.

D) Una investigación sobre la relación entre la estabilidad familiar y la clase social en una ciudad pequeña.

Respuesta

En este caso habría que recurrir como base de la muestra al censo o padrón de vecinos de la ciudad. La unidad de la muestra sería la familia.

E) Afición al cine de horror y de misterio y tipo de personalidad.

Respuesta

En este caso no se ve la posibilidad de obtener como base de la muestra una lista o registro. Por ello habría que adoptar, por ejemplo, el procedimiento de encuestar a uno de cada cinco, ocho, etc., de los asistentes a la proyección de este tipo de películas.

3. En un estudio sociológico sobre una comunidad de 8,500 habitantes mayores de edad se ha elegido una muestra de 850.
Hallar la fracción de muestreo y el coeficiente de elevación. Establecido este último y supuesto que en el estudio realizado han resultado ocho divorciados, determinar el número total de divorciados que es probable exista en la comunidad.

Respuesta

Fracción de muestreo: $\frac{850 \times 100}{8,500} = 10\%$

Coeficiente de elevación: $\frac{8,500}{850} = 10$

Número probable de divorciados en la población: igual al número de divorciados en la muestra por el coeficiente de elevación: 8 x 10 = 80.

4. Se ha proyectado realizar una encuesta sobre el consumo de leche en las familias. El número de familias del universo es 6,000, y la amplitud de la muestra, 840.
Teniendo en cuenta la fracción de muestreo que representa esta muestra, se pide determinar el número de unidades que se deben incluir en la muestra de cada uno de los estratos profesionales siguientes que comprende el universo:

Profesionales liberales	100
Comerciantes	200
Obreros	2,000
Agricultores	600
Servicios	1,900
Empleados	1,200.

Respuesta

Fracción de muestreo: $\frac{840 \times 100}{6,000} = 14\%$

Unidades de muestra para cada estrato:

Profesiones liberales	.....	100.	14 % =	14
Comerciantes	.....	200.	14 % =	28
Obreros	.....	2.000.	14 % =	280
Agricultores	.....	600.	14 % =	84
Servicios	.....	1.900.	14 % =	266
Empleados	.....	1.200.	14 % =	168
TOTAL	.....			840

5. Se ha proyectado realizar una encuesta entre abogados oficialmente en ejercicio de Madrid, con objeto de conocer su opinión sobre la reforma del Código Civil. Señalar las posibles unidades y bases de la muestra, en el caso de que se estimen pueda ser varias, e indicar razonadamente cuál sería más adecuada a los fines perseguidos.

Respuesta

La unidad de la muestra ha de ser, según el enunciado, los abogados en ejercicio, pero hay que advertir que éstos actúan bien individualmente, bien agrupados o bien existen abogados, 'destacados' que tienen en su despacho una serie de ayudantes. Por tanto existe la posibilidad de tomar como unidad de la muestra a cada abogado en ejercicio o bien los despachos y bufetes, ya estén formados por uno o varios abogados.

Dado que el objeto de la investigación es conocer la opinión de los abogados en ejercicio y que cada abogado, aunque trabaje en un despacho con otros, es de suponer que tenga su opinión propia, más o menos diferente de la de sus compañeros, parece que la unidad de la muestra a elegir en este caso debería ser cada abogado individual en ejercicio, independientemente de que trabaje con o a las órdenes de otros. Por otra parte, la elección como unidad de la muestra de los despachos plantea problemas más difíciles para formar la base de la muestra.

Determinada de esta forma la unidad de la muestra, para obtener la base de ésta, o lista de todos los abogados en ejercicio de Madrid, se podría recurrir a varios procedimientos, como, por ejemplo, utilizar los registros fiscales o las listas del Colegio de Abogados. Es obvio que es preferible elegir los registros y listas del Colegio de Abogados, en cuanto constituyen el documento oficial de letrados en ejercicio. Nótese que al dirigir la encuesta a los abogados en ejercicio se excluyen por este hecho mismo a los que estén matriculados en el Colegio como abogados sin ejercicio.

6. Se pretende efectuar una encuesta nacional sobre opiniones, actitudes y valores de la juventud española. Según el censo de 1960, el número de jóvenes de quince a veinticuatro años era, en números redondos, de 4,700,000. La fracción de muestreo se ha fijado en 1/1,000.

Según los datos estadísticos que se poseían, la composición por actividades de la juventud española era la siguiente:

Agricultura	15 por 100
Industria	25 por 100
Comercio	20 por 100
Oficinas y servicios	35 por 100
Estudiantes	5 por 100

El equipo que realiza la encuesta cree que reviste interés especial conocer las opiniones y actitudes de la juventud estudiantil, pero resulta que el número de encuestas a realizar a universitarios, según la proporción que les corresponde, es sólo de 235, número que se considera insuficiente al efecto.

Se pide determinar qué se puede hacer para que, sin variar el tamaño de la muestra, los estudiantes estén representados en ella con un peso cuatro veces superior al que les correspondería según la importancia de su estrato en el universo, indicando la composición resultante de la muestra por estratos.

Respuesta

Como se ha indicado en la introducción, es admisible, cuando un estrato no está suficientemente representado en el universo, forzar o aumentar la representación de este estrato en la muestra, disminuyendo en la misma proporción la participación en ella de otros estratos suficientemente representados.

En este caso, como se dice en el enunciado que se estima que los estudiantes deben estar representados con un peso cuatro veces superior al que les corresponde según el universo, 5 por 100, es claro que su proporción en la

muestra deberá ser de un 20 por 100. Esto supone un aumento del 15 por 100, que deberá ser reducido en otros estratos, por ejemplo, un 8 por 100 en oficinas y servicios, un 4 por 100 en comercio y un 3 por 100 en la industria.

Según esto, la composición de la muestra por estratos sería: %

Agricultura	……	15
Industria	……	22
Comercio	……	16
Servicios y oficinas	……	27
Estudiantes	……	20
TOTAL		**100**

7A Efectuada la encuesta a que se refiere el ejercicio anterior y recogidos los cuestionarios dentro de cada estrato en la proporción prevista, se han declarado partidarios de un sistema económico socialista los siguientes porcentajes –supuestos– dentro de cada estrato.

	%
Agricultura	40
Industria	70
Comercio	40
Servicios	50
Estudiantes	90
Total nacional	59,30

Teniendo en cuenta la alteración realizada en la composición de la muestra con relación al universo de indicar si los porcentajes de cada estrato reflejan efectivamente la extensión de esta actitud en su estrato y si el porcentaje total nacional la refleja en relación a toda la juventud.

Respuesta

Los porcentajes de cada estrato reflejan efectivamente la adhesión a la actitud en cuestión dentro de su estrato, pero no así el total nacional. Este la reflejaría si la composición porcentual de los estratos del universo fuera la misma que la de la muestra. Como según el ejercicio anterior se ha alterado voluntariamente la composición de la muestra, esta alteración implica que en concreto el mayor peso concedido a los estudiantes, que precisamente son los que tienen una actitud más favorable al sistema económico socialista, supondrá un aumento del total nacional respecto al que hubiera resultado si la muestra guardara la misma proporción en los estratos que el universo.

7 B Dada la conclusión anterior, se pide determinar el total nacional efectivo del porcentaje de jóvenes partidarios del sistema económico socialista, corrigiendo la desviación de que se trata mediante la ponderación o multiplicación de los resultados de cada estrato por el porcentaje en el universo y dividiendo el producto por 100.

Respuesta

Jóvenes partidarios de un régimen económico socialista:

GRUPOS DE ACTIVIDAD	% de favorables	% del estrato en el universo	% ponderado de los favorables en cada estado
Agricultura	40	15	6
Industria	70	25	17,5
Comercio	40	20	8
Servicios	50	35	17,5
Estudiantes	90	5	4,5
TOTAL NACIONAL	**59,30**		**53,50**

Se ve que el total nacional partidarios del régimen económico socialista, una vez corregida la alteración efectuada en los estratos de la muestra, es el 53,50 por 100, y no el 59,30 por 100, como resultaba en ésta.

Tanto en este caso, de muestras estratificadas alteradas voluntariamente, como en el de que la muestra obtenida después de efectuada la encuesta difiera sensiblemente en algunos estratos de los porcentajes de éstos en el universo, se deben corregir de este modo los resultados totales.

8. En un sondeo sobre las cuestiones socio–económico–políticas que preocupaban a la población belga, se eligió una muestra de manera estratificada, teniendo en cuenta dos cortes o características: a) la región: flamenca, valona y departamento de Bruselas–capital, y b) el tamaño de la comunidad.

Los porcentajes de representación en la muestra de los distintos estratos de cada corte o característica fueron, una vez realizada la encuesta, los siguientes:

	%
Región flamenca	52,20
Región valona	32,40
Bruselas–capital	15,40
	100,00

	%
De menos de 2.000 habitantes	15
De 2 a 5,000	18,2
De 5 a 10,000	16,6
De 10 a 20,000	15,8
De 20 a 50,000	16,7
De 50 a 100,000	8,6
De 100,000 y más	9,1
	100,00

Verificar el equilibrio de la muestra respecto del universo, teniendo en cuenta que la representación en éste de cada estrato era, en la época de la encuesta, la que sigue:

	%
Región flamenca	49,6
Región valona	33,7
Bruselas-capital	16,7
	100,00

	%
Menos de 2,000 habitantes	15,1
De 2 a 5,000	18,5
De 5 a 10,000	17,2
De 10 a 20,000	14,9
De 20 a 50,000	16,7
De 50 a 100,00	8,3
De 100,000 o más	9,3
	100,00

Respuesta

Para verificar el equilibrio pedido es preciso en primer lugar formar un cuadro comparativo de los porcentajes de cada categoría en la muestra y el universo, similar al que sigue:

CATEGORÍAS	% en muestra	% en universo
Región:		
Flamenca	52,2	49,6
Valona	32,4	33,7
Bruselas–capital	15,4	16,7

Tamaño comunidades:		
De menos de 2,000 habitantes	15,0	15,1
De 2 a 5,000	18,2	18,5
De 5 a 10,000	16,6	17,2
De 10 a 20,000	15,8	14,9
De 20 a 50,000	16,7	16,7
De 50 a 100,000	8,6	8,3
De 100,000 y mas	9,1	9,3

La simple inspección ocular muestra que no existen variaciones importantes en las proporciones de cada estrato en la muestra y en el universo o población total. No obstante, para comprobar si estas diferencias son significativas o se deben simplemente al azar, se puede utilizar la prueba de x_c.

CAPÍTULO III

EL PROCESO DE PRUEBA DE HIPÓTESIS

EL PROCESO DE PRUEBA DE HIPÓTESIS

1. HIPÓTESIS NULAS Y PRUEBA INVERSA

El método científico, en su secuencia final, plantea la necesidad de someter a prueba las hipótesis. Esta tarea es la que hace distinto al conocimiento científico de los otros tipos de conocimientos. Someter a prueba las hipótesis consiste en recolectar datos de la realidad para disponer de evidencia empírica que confirme o contradiga la hipótesis planteada. Para someter a prueba o contrastar una hipótesis es necesario, además de formular la hipótesis alterna, elaborar una hipótesis nula, que viene a ser la negación de la alterna. Es preciso realizar este artificio debido a que es la única manera posible de probar una hipótesis. Esto se debe a que, teóricamente, las hipótesis siempre plantean diferencias, incrementos, efectos, consecuencias, etc., que producirían las variables independientes en las dependientes, y bien se sabe que tales diferencias o incrementos pueden presentarse en diferentes formas, en distintas manifestaciones o en diferentes cantidades. Las hipótesis alternas plantean infinitas maneras de establecer relaciones entre las variables. Por el contrario, las hipótesis nulas, al negar lo planteado en las hipótesis alternas, sostienen que las variables independientes no influyen o no producen ningún efecto en las variables dependientes. Las hipótesis nulas niegan los efectos, niegan las diferencias, niegan los incrementos y tratan de establecer igualdades o semejanzas. En las hipótesis nulas la posibilidad de que una variable no produzca efectos en otra, es una y sólo una, por lo que es más fácil adoptar decisiones con respecto a una hipótesis nula, la que niega los efectos, la que constituye una sola posibilidad, que adoptar decisiones acerca de las diferencias que plantea la hipótesis alterna, que suponen infinitas posibilidades de influencia o incremento. Por ejemplo, dadas las líneas AB y CD, sean las siguientes hipótesis:

A _____ B

C _____ D

Gráfico N° 6

Hipótesis alterna: "*Las líneas son diferentes*"
Hipótesis nula: "*Las líneas no son diferentes*"

Como se ha dicho, existen muchas maneras de que estas líneas sean diferentes, pero existe una sola posibilidad de que estas líneas no sean diferentes, es decir, que sean iguales. La prueba de hipótesis consiste en reunir evidencia empírica con respecto a la hipótesis nula, es decir, reunir datos que pongan en evidencia la única posibilidad en la que las líneas no sean diferentes, que permitan establecer que las líneas sean iguales. Si el investigador acumula datos en apoyo de la hipótesis de la no existencia de diferencias entre las líneas, entonces se acepta la hipótesis nula y se rechaza la hipótesis alterna. Por el contrario, si los datos no confirman la igualdad de las líneas, se rechaza la hipótesis nula y se tiene que aceptar la hipótesis alterna. Pero, en este caso, aceptar la hipótesis alterna no significa comprobar la verdad de tal hipótesis, sólo significa que se ha encontrado evidencia empírica que hace pensar que tal hipótesis no es falsa.

Si un científico sostiene que *la inteligencia incrementa el aprendizaje*, para contrastar esta hipótesis puede formar dos grupos de estudiantes. En uno de ellos puede ubicar estudiantes con alto nivel de inteligencia y en el otro grupo puede ubicar estudiantes con bajo nivel de inteligencia. La teoría científica al respecto conduce a pensar que el grupo de estudiantes que ostenta altos niveles de inteligencia, alcanzará más altos niveles de aprendizaje, mientras que el grupo de estudiantes con bajos niveles de inteligencia alcanzará niveles bajos de aprendizaje. Las diferencias en los niveles de aprendizaje que se puede encontrar entre estos dos grupos son múltiples; se puede hallar pocas diferencias, que las diferencias pueden ser profundas, las diferencias pueden ser de un nivel medio, etc. Existen, en teoría, infinitas posibilidades de que los grupos sean diferentes en cuanto a sus niveles de aprendizaje. Pero si el investigador, cada vez que realiza el experimento, halla las mismas puntuaciones, tanto en grupos de estudiantes con altos niveles intelectuales como en grupos de estudiantes con bajos niveles intelectuales; por ejemplo, si los promedios hallados en ambos grupos fuese, 12,35, entonces podrá decir que acepta la hipótesis nula porque halla, cada vez que examina el aprendizaje en grupos de estudiantes con diferentes niveles intelectuales, promedios de 12,35. Pero esto es un absurdo. No es posible que los grupos de estudiantes con diferentes niveles de inteligencia alcancen puntuaciones iguales, pues las bases teóricas que sustentan esta hipótesis establecen que la inteligencia incrementa el aprendizaje. Sin embargo, si sucediera esta absurda situación, el investigador deberá aceptar la hipótesis nula, aquella que sostiene que no existen diferencias en el aprendizaje entre los grupos de estudiantes con altos niveles de inteligencia con respecto a aquellos grupos de estudiantes que poseen bajos niveles de inteligencia. Ésta constituiría una única situación y al hallarla, debe rechazar la hipótesis alterna, aquella que sostiene que

existen diferencias en los niveles de aprendizaje entre ambos grupos de estudiantes. Pero como ésta es la situación absurda, en los hechos, cabe la posibilidad de que el investigador se incline por rechazar la hipótesis nula, pues es más posible encontrar diferencias que igualdades.

Esta es la manera de contrastar hipótesis. La metodología científica indica que se debe proceder de este modo, pues es más fácil hallar la única posibilidad que no se produzcan diferencias que demostrar las infinitas posibilidades de diferencias que plantean las hipótesis alternas. Algunos autores llaman a este procedimiento la prueba del absurdo, pues se asume, como hipótesis, una situación absurda para luego buscar evidencias que nieguen o contradigan esta absurda situación.

2. ERRORES AL ADOPTAR DECISIONES

En base a la evidencia empírica reunida, el investigador adopta sus decisiones y, como es natural, toda decisión humana lleva implícita la posibilidad de error. El investigador debe tratar de no cometerlos; sin embargo, cuando se adoptan decisiones con respecto a la hipótesis, el investigador puede cometer dos tipos de errores:

Error de tipo I: El error de tipo I consiste en aceptar la hipótesis nula cuando en realidad es falsa. Para evitarlo, se recomienda ampliar el tamaño de la muestra.

Error de tipo II: El error de tipo II consiste en rechazar la hipótesis nula cuando en realidad no existen argumentos para rechazarla. Para evitarlo se recomienda perfeccionar la calidad de los instrumentos de acopio de datos.

Considerando este contexto, el investigador siempre correrá el riesgo de cometer errores, por lo que debe trazar estrategias más eficaces, utilizar instrumentos más precisos y trabajar con muestras del tamaño adecuado.

Es normal que se produzcan estos errores. El avance científico tecnológico depende, en muchos casos, de las correcciones que hacen los científicos a los errores que han cometido otros científicos que realizaron investigaciones precedentes. Los esfuerzos por realizar replicaciones o refutaciones científicas permiten corregir los errores que un investigador hubiera cometido o confirmar las decisiones a las que puede haber llegado luego de repetir los experimentos en varias oportunidades, lo que permite confirmar la validez de las decisiones adoptadas.

3. PRUEBA DE HIPÓTESIS

El proceso de prueba de hipótesis puede hacerse de dos maneras: mediante técnicas estadísticas o mediante técnicas que no requieren el empleo de la estadística. Se ha dado en llamar investigación cuantitativa cuando se usan técnicas estadísticas y se denomina investigación cualitativa cuando no se usan técnicas estadísticas. En la actualidad se desarrolla una polémica muy intensa en torno a la validez de estas técnicas, pero parece ser que algunas variables, necesariamente deben ser estudiadas con métodos cuantitativos, porque es posible medir o cuantificar sus magnitudes o propiedades, mientras que otras variables, por su propia naturaleza, no pueden estimarse cuantitativamente, entonces se deben emplear, necesariamente, los métodos cualitativos. En muchos casos, las últimas investigaciones científico sociales se realizan integrando el análisis cuantitativo con el cualitativo, lo que permite una mejor comprensión de los fenómenos que se estudian. En lo que sigue, se explicará la metodología que emplea técnicas estadísticas.

Cuando se emplean métodos estadísticos o cuantitativos, se dispone de dos tipos de pruebas estadísticas: las paramétricas y las no paramétricas. Las pruebas paramétricas sirven para analizar eventos producidos por el azar o la suerte, mientras que las pruebas no paramétricas sirven para analizar eventos producidos por la intención, la voluntad o los propósitos. En la investigación de la conducta, campo en el que predomina la voluntad y la intención de los individuos, las hipótesis se contrastan con pruebas no paramétricas. En cambio, las pruebas paramétricas se usan cuando las hipótesis tienen que ver con situaciones en las que no actúa la voluntad o la intención de las personas, es decir, cuando los hechos se producen al azar. Es más plausible usar pruebas paramétricas en la investigación en ciencias naturales.

4. NIVEL DE SIGNIFICACIÓN

En todo proceso de investigación cabe suponer que los eventos pueden ocurrir por causa del azar o de la intención. El problema radica en determinar cuándo un evento se produce por causas del azar o cuándo un evento se produce por causas de la intención. El nivel de significación proporciona los criterios para decidir acerca de esta situación.

El nivel de significación es el margen de tolerancia aceptable para establecer los límites, dentro de los cuales, se debe decidir si los eventos ocurren por causas del azar o por causas de la intención. El nivel de significación se expresa en términos de porcentajes. Para el caso de la investigación de la conducta el porcentaje aceptable es del 5%, lo que significa que para aceptar una hipótesis alterna tendrá que ser necesario que los eventos sucedan en el 95% de los casos, y sólo en el 5% sucedan por causas del azar. Por ejemplo, si el investigador postula la siguiente hipótesis: *el*

método didáctico de trabajo en equipo permite mejorar el nivel de aprendizaje de los estudiantes, y encuentra que en el grupo en el que ha aplicado la variable *método didáctico de trabajo en equipo*, el nivel de aprendizaje del grupo, en promedio, es 13,45, mientras que en el grupo en el que no ha aplicado dicha variable el promedio del grupo llega a 13,06, debe decidir si la diferencia entre estas dos puntuaciones, que sólo es de 0,39 centésimos, ha sido producida por la aplicación de la variable *método didáctico de trabajo en equipo* o simplemente ha sido producida por la casualidad o el azar. En estas situaciones, el investigador no tiene elementos de juicio para decidir si los hechos se han producido por el azar o por la aplicación de la variable, que es la situación intencionada. Como es natural, si las diferencias que halla son mínimas, cabe suponer que los hechos se producen por la casualidad o el azar. En cambio, si las diferencias son notorias, cabe suponer que los hechos se han producido como consecuencia de la aplicación de la variable. El problema que tiene el investigador, en esta etapa del proceso, radica en establecer a qué llama pequeña diferencia o diferencia no significativa o diferencia grande o significativa. El nivel de significación es el criterio que le permite adoptar la decisión más adecuada.

El nivel de significación también se puede expresar en términos decimales: En este caso, 5% equivale a 0,05 ya que el margen de confianza es 0,95. La suma de estas cifras da 1. En las Ciencias Sociales, el nivel de significación que generalmente se usa es de 0,05. En otras disciplinas científicas se usan niveles de significación más finos, es decir, se acepta que los hechos ocurran al azar en porcentajes menores como por ejemplo: el 0,5%, 0,1% ó 0,01%, que en términos decimales, estos niveles de significación son del 0,005, 0,001 y 0,0001, respectivamente.

En cambio, en los estudios de opinión o de mercado, debido a que la voluntad de los sujetos es muy cambiante, se aceptan márgenes de error o niveles de significación más amplios, como por ejemplo del 10%, es decir, 0,10.

5. GRADOS DE LIBERTAD

Los grados de libertad son las posibilidades de libre variación que tienen las categorías de una variable. Si por ejemplo se debe expresar las categorías de variación de la variable *rendimiento académico* y se tienen los siguientes datos: *matriculados* 48, *aprobados* 41, los *desaprobados* ¿cuántos serán? Estos serán necesariamente 7, que es la diferencia entre estas dos cifras propuestas al azar. En este ejemplo existen tres categorías –matriculados, aprobados y desaprobados– y los grados de libertad son dos, porque 48 y 41 son cifras arbitrarias y 7 es la cifra que expresa la diferencia entre aquellas. Los grados de libertad se calculan aplicando la siguiente fórmula:

$$gl = K–1$$

en este caso, gl = 3–1 = 2. Lo que significa que dadas tres categorías de variación, los grados de libertad son 2.

Tanto el establecimiento del nivel de significación como la identificación de los grados de libertad son tareas necesarias para poder interpretar los datos que se obtienen en el proceso de prueba de hipótesis. En la investigación cuantitativa, al realizar el proceso de prueba de hipótesis, se obtiene un dato, una cifra, a la que se denomina valor hallado que no tiene ningún sentido si no es comparada con otra cifra, llamada valor tabulado. Este valor tabulado, aparece en unas tablas que generalmente se publican como anexos de los libros de estadística inferencial. La tarea del investigador consiste en comparar el valor hallado, que es el que ha encontrado luego de hacer los respectivos cálculos, con el valor tabulado. En las tablas, se puede ubicar el valor tabulado identificando la intersección entre el nivel de significación y los grados de libertad previstos. De ahí que resultan muy importantes para la toma de decisiones con respecto a las hipótesis, los conceptos de nivel de significación y grados de libertad.

6. PRUEBA DE CHI CUADRADA (X^2)

Una de las pruebas no paramétricas más conocidas es la prueba de Chi cuadrada. Esta prueba ayuda a adoptar la decisión más adecuada con respecto a la hipótesis. Sus fórmulas tienen algunas variaciones según los datos que se desee estudiar, pero en todos los casos se requiere que los datos se presenten en forma de frecuencias.

Sea la siguiente investigación: Estudio de las actitudes de los profesores con respecto al ascenso por méritos. La hipótesis sería: *Los profesores, a medida que avanzan en edad, cambian de actitud con respecto al ascenso por méritos*. Esto significa, hipotéticamente, que los profesores jóvenes están de acuerdo con el ascenso por méritos en la carrera profesional, mientras que los profesores de más edad, están en contra de este tipo de criterio para el ascenso. Sean, por ejemplo, los siguientes datos:

EDAD	FAVORABLE	NEUTRAL	DESFAVORABLE	TOTAL
21 – 25	15 (8,6)	2 (3,6)	3 (7,8)	20
26 – 30	11 (8,6)	4 (3,6)	5 (7,8)	20
31 – 35	9 (8,6)	5 (3,6)	6 (7,8)	20
36 – 40	5 (8,6)	5 (3,6)	10 (7,8)	20
41 – 45	3 (8,6)	2 (3,6)	15 (7,8)	20
N	43	18	39	100

Cuadro Nº 5

Los datos se presentan en una tabla de contingencia 3 x 5, por lo que los grados de libertad son 8.

Para el cálculo de los grados de libertad se ha procedido de la siguiente manera:

$$gl = (K - 1)(L - 1)$$

"K", representa el número de columnas que tiene la tabla de contingencia. En este caso, las columnas son 3, que corresponden a las actitudes favorable, neutral y desfavorable.

"L" representa el número de líneas que tiene la tabla de contingencia. En este caso, las líneas son 5, que corresponden a los cinco grupos etáreos en que se han agrupados a los sujetos: 21 a 25 años, 26 a 30, 31 a 35, 36 a 40 y 41 a 45.

Con estos datos, deben hacerse las siguientes operaciones:

$$gl = (3 - 1)(5 - 1) = 2 \times 4 = 8$$

En este caso, las decisiones se adoptarán con 8 grados de libertad y al nivel de significación de 0,05

La fórmula de Chi cuadrada que se empleará en este caso es la siguiente:

$$X^2 = \Sigma \frac{(fo - fe)^2}{fe}$$

En donde: Σ = Sumatoria.
fo = frecuencia obtenida.
fe = frecuencia esperada.

Como la fórmula exige hallar la sumatoria total del cuadrado de las diferencias entre la frecuencia obtenida y la frecuencia esperada, dividida entre la frecuencia obtenida, se requiere hallar el cuadrado de las diferencias entre la frecuencia obtenida y la frecuencia esperada y dividirla entre la frecuencia esperada que corresponde a cada una de las tres actitudes. Para ello es necesario organizar los datos de la siguiente manera.

DATOS PARA EL CÁLCULO
DE CHI CUADRADA DE LA ACTITUD FAVORABLE

Edades	fo	fe	fo-fe	$(fo-fe)^2$	$\frac{(fo-fe)^2}{fe}$
21-25	15	8,6	6,4	40,96	4,76
26-30	11	8,6	2,4	5,76	0,66
31-35	9	8,6	0,4	0,16	0,0018
36-40	5	8,6	-3,6	12,96	1,51
41-45	3	8,6	-5,6	31,36	3,65
N	43	43			10,5818

Cuadro Nº 6

DATOS PARA EL CÁLCULO
DE CHI CUADRADA DE LA ACTITUD NEUTRAL

Edades	fo	fe	fo-fe	$(fo-fe)^2$	$\frac{(fo-fe)^2}{fe}$
21-25	2	3,6	-1,6	2,56	0,71
26-30	4	3,6	0,4	0,16	0,04
31-35	5	3,6	1,4	1,96	0,54
36-40	5	3,6	1,4	1,96	0,54
41-45	2	3,6	-1,6	2,56	0,71
N	18	18			2,54

Cuadro Nº 7

DATOS PARA EL CÁLCULO
DE CHI CUADRADA DE LA ACTITUD DESFAVORABLE

Edades	fo	fe	fo-fe	$(fo-fe)^2$	$\frac{(fo-fe)^2}{fe}$
21-25	3	7,8	-4,8	23,04	2,95
26-30	5	7,8	-2,8	7,84	1,00
31-35	6	7,8	-1,8	3,24	0,42
36-40	10	7,8	2,2	4,84	0,62
41-45	15	7,8	7,2	51,84	6,65
N	39	39			11,64

Cuadro Nº 8

Las sumatorias de Chi cuadrada obtenidas en cada caso son las siguientes, tal como pueden verse en los respectivos cuadros:

Para la actitud favorable: 10,58
Para la actitud neutral: 2,54
Para la actitud desfavorable: 11,64

La sumatoria total es: 24,76. Este es el valor de Chi cuadrada hallado o encontrado para estos datos. Seguidamente se recurre a la Tabla de Valores Críticos de Chi cuadrada y se ubica el valor que se encuentre en la intersección de la fila que corresponde a 8 grados de libertad y de la columna que corresponde al nivel de significación 0,05, para una prueba bilateral. Se dice que la prueba o la hipótesis es bilateral cuando no señala el sentido de las diferencias, es decir, no especifica si la variable independiente incrementa o disminuye el valor de la variable dependiente,

sólo plantea la influencia; mientras que una prueba o una hipótesis unilateral sí es específica, expresa que la variable independiente incrementa o, por el contrario, disminuye, los valores de la variable dependiente.

En el presente caso, el valor hallado en la tabla es 15,51, tal como puede verse en el cuadro N° 9. En este caso, como el valor encontrado: 24,76 es mayor que el valor que aparece en la tabla, se debe adoptar la decisión de rechazar la hipótesis nula que sostiene: *Los profesores no cambian de actitud con respecto al ascenso por méritos a medida que aumentan en edad* y se debe aceptar la hipótesis alterna que sostiene: *Los profesores cambian de actitud con respecto al ascenso por méritos a medida que aumentan en edad.*

VALORES CRÍTICOS DE CHI CUADRADA

Gl	Nivel de significación para una prueba bilateral					
	0,20	**0,10**	**0,05**	**0,02**	**0,01**	**0,001**
1	1,64	2,71	3,84	5,41	6,64	10,87
2	3,22	4,60	5,99	7,82	9,21	13,82
3	4,64	6,25	7,82	9,84	11,34	16,27
4	5,99	7,78	9,49	11,67	13,28	18,46
5	7,29	9,24	11,07	13,39	15,09	20,52
6	8,56	10,64	12,59	15,03	16,81	22,46
7	9,80	12,02	14,07	16,62	18,48	24,32
8	11,03	13,36	15,51	18,17	20,09	26,12
9	12,24	14,68	16,92	19,68	21,67	27,88
10	13,44	15,99	18,31	21,16	23,21	29,59
11	14,63	17,28	19,68	22,62	24,72	31,26
12	15,81	18,55	21,03	24,05	26,22	32,91
13	16,98	19,81	22,36	25,47	27,69	34,53
14	18,25	21,06	23,68	26,87	29,14	36,12
15	19,31	22,31	25,00	28,26	30,58	37,70
16	20,46	23,54	26,30	29,63	32,00	39,20
17	21,62	24,77	27,59	31,00	33,41	40,75
18	22,76	25,99	28,87	32,35	34,80	42,31
19	23,90	27,20	30,14	33,69	36,19	43,82
20	25,04	28,41	31,41	35,02	37,57	45,32
21	26,17	29,62	32,67	36,34	38,93	46,80

Cuadro N° 9

7. ANÁLISIS DE VARIANZA

Cuando en la investigación se aplica una estrategia factorial o multivariada, es decir, cuando intervienen dos o más variables independientes y éstas no se encuentran correlacionadas entre sí, se recomienda aplicar el Análisis de Varianza como método más apropiado para probar hipótesis.

Para aplicar el análisis de varianza se debe, en primer lugar, trazar una estrategia adecuada. Si por ejemplo el investigador estudia los efectos que dos variables independientes producen en una dependiente y cada una de estas variables asume dos valores, la estrategia consistirá en acomodar los datos en una tabla de contingencia 2x2, que algunos autores llaman rejilla 2x2 ó tabla de divisiones cruzadas 2x2. 2x2 significa que se analizarán los efectos de dos variables independientes cuando cada una de ellas varía en dos valores.

Al aplicar el análisis de varianza, y para mantener una misma nomenclatura, a cada una de las variables se denomina factor. Una de ellas es el factor A y la otra es el factor B. Pero como cada variable asume dos valores, el factor A variará en A_1 y A_2 y el factor B variará en B_1 y B_2. El sub índice 1 significa que la variable se presenta en su nivel más bajo de variación y el sub índice 2 significa que la variable se presenta en su nivel más alto de variación.

Sea la siguiente tabla de contingencia:

		Factor A: A_1	Factor A: A_2
Factor B:	B_1	*[1]* $A_1 A_1$	*[a]* $A_2 B_1$
	B_2	*[b]* $A_1 B_2$	*[ab]* $A_2 B_2$

Gráfico N° 7

Como resultado de acomodar las variables en la tabla de contingencia 2x2 se produce un cruce de variables, y como consecuencia de ello es posible distinguir los siguientes cuatro tratamientos: A_1, B_1; A_2, B_1; A_1, B_2; A_2, B_2 .

El ejemplo que ilustra lo dicho es el siguiente: Suponiendo que se desea estudiar los efectos de las variables *desempeño docente* y *métodos didácticos*. El *desempeño docente* es el factor A y varía en A_1, que corresponde a un *desempeño*

docente no eficiente y A_2 que corresponde a un *desempeño docente eficiente*. Como se aprecia, el desempeño docente no eficiente es la presencia de la variable en su nivel más bajo, no eficiente, o sea *1*. Mientras que A_2, corresponde al *desempeño docente eficiente*, es decir, el valor alto de la variable: *2*. Por otra parte, el factor *métodos didácticos* puede variar también en dos niveles: B_1, aplicación de *métodos didácticos centrados en la enseñanza* o valor bajo de la variable B_1, y B_2, aplicación de *métodos didácticos centrados en el aprendizaje*, o valor alto de la variable B. La teoría pedagógica sostiene que el problema fundamental en todo proceso educativo es el aprendizaje y no la enseñanza, de ahí que se considera *valor bajo* a los *métodos didácticos centrados en la enseñanza* y *valor alto* a los *métodos didácticos centrados en el aprendizaje*.

La hipótesis que se puede elaborar con estos datos es la siguiente: *El eficiente desempeño docente y los métodos didácticos centrados en el aprendizaje son factores que contribuyen a incrementar el nivel de rendimiento académico de los estudiantes*.

Con estos datos, la primera combinación de las variables es A_1, B_1, es decir grupo de alumnos que tiene *profesores con desempeño docente no eficiente* y a quienes se les aplica *métodos didácticos centrados en la enseñanza*.

El primer tratamiento es, entonces, el tratamiento A_1, B_1, en el que los valores de las variables se presentan en sus niveles más bajos. Ambos factores están en *1* y para simplificar la denominación de este tratamiento se asume la convención de llamarlo tratamiento *[1]*, que en realidad significa el tratamiento menos deseado, el tratamiento en el que prácticamente no actúan o no influyen las variables. A partir de este tratamiento se puede elaborar la hipótesis nula.

El segundo tratamiento, o segundo grupo, corresponde a la combinación A_2, B_1, es decir el grupo de estudiantes que tiene *profesores con desempeño docente eficiente* y a quienes se les aplica *métodos didácticos centrados en la enseñanza*. En este grupo o tratamiento actúa la variable *A*, es decir, la variable está en su valor más alto: *desempeño docente eficiente*. Para simplificar su denominación a este tratamiento se le conoce como tratamiento *[a]*, porque A, y no B, asume el valor alto; B permanece en el nivel bajo.

El tercer campo del cuadro es el tratamiento A_1, B_2, es decir, el grupo de alumnos que tiene *profesores con desempeño docente no eficiente* pero a quienes se les aplica *métodos didácticos centrados en el aprendizaje*. En este grupo o tratamiento actúa la variable B, es decir, la variable está en su valor más alto: *métodos didácticos centrados en el aprendizaje*, B_2. Para simplificar su denominación a este tratamiento se le conoce como tratamiento *[b]*, porque B asume el valor alto, mientras que la variable A no actúa.

El cuarto campo del cuadro es el tratamiento A_2, B_2, es decir el grupo de estudiantes que tiene *profesores con desempeño docente eficiente* y a quienes se les aplica *métodos didácticos centrados en el aprendizaje*. En este grupo o tratamiento actúan las variables A y B, es decir ambas están en sus valores más altos: *profesores con desempeño didáctico eficiente*, A_2, y *métodos didácticos centrados en el aprendizaje*, B_2. Para simplificar su denominación a este tratamiento se le conoce como tratamiento *[ab]*, porque A y B asumen sus valores más altos.

Como se ha dicho, el análisis de varianza es una metodología que se aplica para probar hipótesis siempre y cuando las variables independientes no se encuentren relacionadas entre sí. Algunas veces sucede que las variables independientes que actúan sobre la dependiente se hallan, a su vez, relacionadas entre sí, o dicho en otros términos, existe covarianza entre las variables independientes. Cuando la covarianza entre las variables independientes no existe, o es igual a cero, se puede aplicar el análisis de varianza.

Cuando se presenta esta situación, el investigador puede asumir que las variables influyen por separado, cada una en su oportunidad, en la variable dependiente. Por eso es que se puede hacer el análisis, en forma independiente, de cada uno de los factores, y por eso es que se puede elaborar la estrategia de formar grupos o tratamientos que se ubican en una tabla de contingencia como la descrita. En el ejemplo que se está presentando, existen dos variables y cada una de ellas asume dos valores, de ahí que la tabla de contingencia, como se ha dicho, se denomina 2x2, o sea, existe la posibilidad de formar cuatro grupos. Si una de las variables asumiera tres valores, la tabla se denominaría 2x3, o sea se formarán 6 grupos o tratamientos. Las tablas de contingencia se tornan cada vez más complejas, es decir, aparecen más grupos de tratamiento a medida que las variables asuman más valores o en la hipótesis se consideren tres o más variables independientes.

En el ejemplo que se está presentando, existen dos variables y cada una de ellas asume dos valores, de ahí que resulta la posibilidad, o contingencia, de formar cuatro grupos a quienes se puede aplicar distintos tratamientos. Los valores hallados para la variable dependiente se ubican en cada celda o recuadro de la tabla de contingencia 2x2.

El análisis de varianza consiste en analizar la varianza de las puntuaciones alcanzadas por los cuatro grupos. Este análisis se hace desde dos perspectivas, la primera consiste en analizar la varianza que podría existir al interior de cada uno de los grupos, es la varianza dentro de los grupos, o varianza intra grupo, de ahí las denominaciones que se dan a los datos que se encuentran: suma de cuadrados dentro de los grupos, SS_D, o media de cuadrados dentro de los grupos, MS_D, en

donde el sub índice D expresa que el análisis se ha hecho con las puntaciones obtenidas por los sujetos dentro de cada grupo.

La segunda perspectiva es realizar el análisis de la varianza entre cada uno de los grupos, de ahí las denominaciones de los datos que se encuentran: suma de cuadrados entre grupos, SS_E, o media de cuadrados entre grupos, MS_E, en donde el sub índice E expresa que el análisis se ha hecho comparando las puntaciones obtenidas por los sujetos entre cada uno de los grupos.

El siguiente gráfico expresa lo dicho hasta el momento:

		Factor A: *Desempeño docente*	
		A_1 *No eficiente*	A_2 *Eficiente*
Factor B: *Métodos didácticos*	B_1 *Centrados en la enseñanza*	*[1]* $A_1 A_1$	*[a]* $A_2 B_1$
	B_2 *Centrados en el aprendizaje*	*[b]* $A_1 B_2$	*[ab]* $A_2 B_2$

Gráfico N° 8

A manera de ejemplos, se presentan a continuación la síntesis de algunos trabajos realizados por investigadores de la Facultad de Educación de la U.N.M.S.M. El que sigue es el estudio realizado por Constantino Domínguez Barrera, titulado *Influencia del Desempeño Docente y de los Métodos Didácticos en el Rendimiento Académico de Alumnos del Ciclo Básico de Obstetricia*. En este caso se investiga la influencia de dos variables independientes sobre una variable dependiente.

Problema

¿Qué efectos producen, en el rendimiento académico de alumnos de Ciclo Básico de Obstetricia, el desempeño docente y los métodos didácticos que se emplean en el proceso enseñanza-aprendizaje?

Variación de las variables:

Desempeño docente, asume dos valores: *Desempeño docente no eficiente*, o nivel bajo de la variable, (A_1), y *Desempeño docente eficiente*, o nivel alto de la variable, (A_2).

Métodos didácticos: *Métodos didácticos centrados en la enseñanza*, o nivel bajo de la variable, (B_1), y *Métodos didácticos centrados en el aprendizaje*, o nivel alto de la variable, (B_2).

Sistema de hipótesis

Hipótesis general

El eficiente desempeño docente, (A_2), y la aplicación de métodos didácticos centrados en el aprendizaje, (B_2), incrementan significativamente el nivel de rendimiento académico de estudiantes del Ciclo Básico de Obstetricia.

En términos formales: *el rendimiento académico, representado por el símbolo Y es influenciado, en forma positiva, por el desempeño docente eficiente* (A_2) *y por la aplicación de métodos didácticos centrados en el aprendizaje* (B_2).

Es decir, se plantea el siguiente modelo estadístico propuesto por Jaime Arnau:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \in_{ijk}$$

En donde:

Y_{ijk} es la 'k-ésima' observación bajo el 'i-ésimo' tratamiento de la variable A (Desempeño docente) y bajo el 'j-ésimo' tratamiento de la variable B (Métodos didácticos). (i = 1,2; j = 1,2; k = 1,2,3 ... 80)

μ :Es la media general del experimento y recoge todos los efectos constantes.

α_i: Es el efecto del i-ésimo nivel del factor A (Desempeño docente).

β_j : Es el efecto del j-ésimo nivel del factor B (Métodos didácticos).

$(\alpha\beta)_{ij}$: Es el efecto de la interacción de las variables.

$\in$: Es el error estimado.

Este modelo de hipótesis permite estimar los efectos de cada una de las variables independientes en el *rendimiento académico*.

Sub hipótesis alterna 1

El rendimiento académico del grupo de estudiantes del Ciclo Básico de Obstetricia, que tuvo profesores con desempeño docente eficiente, (A_2)*, es mayor que el rendimiento académico del grupo de estudiantes que tuvo*

profesores con desempeño docente no eficiente, (A_1). En términos formales, para el factor A:

$$H_1 : \mu_{20} > \mu_{10}$$

Sub hipótesis nula 1

El rendimiento académico del grupo de estudiantes del Ciclo Básico de Obstetricia, que tuvo profesores con desempeño docente eficiente, (A_2), *no difiere significativamente del rendimiento académico del grupo de estudiantes que tuvo profesores con desempeño docente no eficiente,* (A_1). En términos formales, para el factor A:

$$H_1 : \mu_{20} = \mu_{10}$$

Sub hipótesis alterna 2

El rendimiento académico del grupo de estudiantes del Ciclo Básico de Obstetricia, a quienes se les aplicó métodos didácticos centrados en el aprendizaje, (B_2), *es mayor que el rendimiento académico del grupo de estudiantes a quienes se les aplicó métodos didácticos centrados en la enseñanza,* (A_1). En términos formales, para el factor B:

$$H_1 : \mu_{02} > \mu_{01}$$

Sub hipótesis nula 2

El rendimiento académico del grupo de estudiantes del Ciclo Básico de Obstetricia, a quienes se les aplicó métodos didácticos centrados en el aprendizaje, (B_2), *no difiere significativamente del nivel de rendimiento académico del grupo de estudiantes a quienes se les aplicó métodos didácticos centrados en la enseñanza,* (A_1). En términos formales, para el factor B:

$$H_1 : \mu_{02} = \mu_{01}$$

Sub hipótesis alterna 3

La interacción de los factores desempeño docente y métodos didácticos, (A x B), *produce efectos significativos en el rendimiento académico de estudiantes del Ciclo Básico de Obstetricia*. En términos formales:

$$H_3 : (\alpha\beta)_{ij} > 0 \text{ para todo ij}$$

Sub hipótesis nula 3

La interacción de los factores desempeño docente y métodos didácticos, (A x B), no produce efectos significativos en el rendimiento académico de estudiantes del Ciclo Básico de Obstetricia. En términos formales:

$$H_3 : (\alpha\beta)_{ij} = 0 \text{ para todo ij}$$

Proceso de prueba de hipótesis. Método I

Luego de la aplicación de los instrumentos de acopio de datos, en este caso el test para medir el rendimiento académico de los estudiantes de la muestra, los datos encontrados fueron los siguientes:

		Factor A: *Desempeño docente*	
		A_1 *No eficiente*	A_2 *Eficiente*
Factor B: *Métodos didácticos*	B_1 *Centrados en la enseñanza*	*[1]* 867	*[a]* 998
	B_2 *Centrados en el aprendizaje*	*[b]* 985	*[ab]* 1087

Gráfico N° 9

CÁLCULO DE LOS EFECTOS DE LOS FACTORES MEDIANTE LA COMBINACIÓN LINEAL

Factores	Combinación lineal				Total	Efecto de los factores
	1	a	b	ab		
	867	998	985	1087		
Efecto del factor A	-	+	-	+	233	1,46
Efecto del factor B	-	-	+	+	207	1,30
Efecto de AxB	+	-	-	+	-29	-0,18

Cuadro N° 10

Para el cálculo de los efectos de los factores mediante la combinación lineal, se han colocado los datos de la tabla de contingencia 2x2 en la secuencia lineal: 1, a, b, y ab.

Para calcular el efecto del factor *A, desempeño docente*, se suma algebraicamente los valores de cada tratamiento. Se considera que en el tratamiento *[1]*, el factor A no interviene debido a que está en su nivel bajo: A_1, por lo que este

tratamiento está señalado con el signo negativo. Se considera que en el tratamiento *[a]*, el factor A interviene debido a que está en su nivel alto: A_2, por lo que este tratamiento está señalado con el signo positivo. Se considera que en el tratamiento *[b]*, el factor A no interviene debido a que está en su nivel bajo: A_1, por lo que este tratamiento está señalado con el signo negativo. Se considera que en el tratamiento [*ab*], el factor A interviene debido a que está en su nivel alto: A_2, por lo que este tratamiento está señalado con el signo positivo. En consecuencia la suma algebraica para este factor es 233. Este puntaje ha sido alcanzado debido a que hay 80 alumnos en cada uno de los tratamientos signados con el signo positivo. Como para este factor hay dos tratamientos con signo positivo, se considera que son 160 los sujetos que han hecho posible alcanzar los 233 puntos en la suma algebraica; dividiendo 233 entre los 160 sujetos, se obtiene el valor de 1,46 que se considera es el valor con que el factor A, en este caso, *desempeño docente* influye en el *rendimiento académico* de los estudiantes de Obstetricia. Y si el factor A, *desempeño docente* influye en 1,46 en el *rendimiento académico*, se puede adoptar la decisión de rechazar la sub hipótesis nula 1 y aceptar la sub hipótesis alterna 1, que sostiene: *El rendimiento académico del grupo de estudiantes del Ciclo Básico de Obstetricia, que tuvo profesores con desempeño docente eficiente,* (A_2)*, es mayor que el rendimiento académico del grupo de estudiantes que tuvo profesores con desempeño docente no eficiente.*

De manera similar al caso anterior se procede para analizar la influencia del factor B: métodos didácticos. Se considera que en el tratamiento [*1*], el factor B no interviene debido a que está en su nivel bajo: B_1, por lo que este tratamiento está señalado con el signo negativo. Se considera que en el tratamiento [*a*], el factor B no interviene debido a que también está en su nivel bajo: B_1, por lo que este tratamiento está señalado con el signo negativo. Se considera que en el tratamiento [*b*], el factor B interviene debido a que está en su nivel alto: B_2, por lo que este tratamiento está señalado con el signo positivo. Se considera que en el tratamiento [*ab*], el factor B interviene debido a que también está en su nivel alto: B_2, por lo que este tratamiento está señalado con el signo positivo. En consecuencia, la suma algebraica para este factor es 207. Este puntaje ha sido alcanzado debido a que hay 80 alumnos en cada uno de los tratamientos signados con el signo positivo. Como para este factor hay dos tratamientos con signo positivo, se considera que son 160 los sujetos que han hecho posible alcanzar los 207 puntos en la suma algebraica. Dividiendo 207 entre los 160 sujetos, se obtiene el valor de 1,29 que se considera es el valor con que el factor B, en este caso, *métodos didácticos*, influye en el *rendimiento académico* de estudiantes de Obstetricia. Y si el factor B, *métodos didácticos*, influye en 1,29 en el *rendimiento académico*, se puede adoptar la decisión de rechazar la hipótesis nula 2 y aceptar la sub hipótesis alterna 2, que

sostiene: *El rendimiento académico del grupo de estudiantes del Ciclo Básico de Obstetricia, a quienes se les aplicó métodos didácticos centrados en el aprendizaje,* (B_2)*, es mayor que el rendimiento académico del grupo de estudiantes a quienes se les aplicó métodos didácticos centrados en la enseñanza,* (A_1).

Para analizar la interacción o el efecto conjunto de los factores A y B se procede del siguiente modo: En el tratamiento [*1*] los factores A y B están en sus valores más bajos, es decir, A_1B_1, por lo que se puede decir que en este tratamiento ambos factores se encuentran iguales, por lo que se les reconoce con el signo positivo. En el tratamiento [*a*], sólo el factor A está en su nivel alto, por tanto, al no haber interacción, corresponde colocar el signo negativo. En el tratamiento [*b*], sólo el factor B está en su nivel alto, por tanto, también corresponde colocar el signo negativo. En cambio, en el tratamiento [*ab*] ambos factores están en sus niveles altos, A_2B_2, por lo que corresponde colocar el signo positivo. De ahí que la suma algebraica para la interacción o el efecto conjunto de los factores A y B sea –29 que dividido entre 160 resulta -0,18, es decir, un valor menor que cero, por lo que se adopta la decisión de aceptar la sub hipótesis nula 3 que sostiene: *La interacción de los factores desempeño docente y métodos didácticos,* (A x B)*, no produce efectos significativos en el rendimiento académico de estudiantes del Ciclo Básico de Obstetricia,* y rechazar la sub hipótesis alterna 3. Esta decisión concuerda con la teoría al respecto, pues los factores A y B actúan independientemente, no se entremezclan ni se traslapan, no producen efectos conjuntos, de ahí que el valor hallado para la hipótesis de la interacción sea una cifra muy pequeña, incluso menor que cero.

CÁLCULO DE LOS EFECTOS DE LOS FACTORES A PARTIR DE LOS PROMEDIOS

	A_1	A_2	Totales	Medias
B_1	[1] 10,84	[a] 12,48	23,32	$Y_{01} = 11,66$
B_2	[b] 12,31	[ab] 13,59	25,90	$Y_{02} = 12,95$
Totales	23,15	26,07	49,22	
Medias	$Y_{10} = 11,58$	$Y_{20} = 13,04$		$Y_{00} = 12,31$

Cuadro Nº 11

Empleando este método, los resultados son iguales a los encontrados con el método I, pues al hacer las operaciones correspondientes se obtiene:

Para el factor A: Desempeño docente:

$\alpha_1 = Y_1 - Y = 11,58 - 12,31 = -0,73$

$\alpha_2 = Y_2 - Y = 13,04 - 12,31 = 0,73$

Total: –0.73 + 0,73 = 1,46

Para el factor B: Métodos didácticos:

$\beta_1 = Y_{01} - Y = 11,66 - 12,31 = -0,65$

$\beta_2 = Y_{02} - Y = 12,95 - 12,31 = 0,64$

Total: –0,65 + 0.64 = 1,29

Método II. Análisis entre y dentro de los grupos

Como se ha dicho, el análisis de varianza realizado a un nivel más profundo consiste en analizar la varianza al interior de los grupos o varianza dentro de los grupos, y la varianza entre los grupos. Luego de hacer este análisis se obtiene la suma de cuadrados dentro de los grupos y la suma de cuadrados entre los grupos. Los procesadores estadísticos que se disponen últimamente permiten realizar todos estos cálculos que, hechos manualmente, suponen algún nivel de complejidad para quienes no son expertos en temas estadísticos, de modo que en este punto obviamos la descripción de todo este proceso.

Prueba F

Para adoptar las decisiones con respecto a la hipótesis, se aplica la prueba F que arrojará un valor como consecuencia de dividir la media de cuadrados entre grupos, sobre la media de cuadrados dentro de los grupos. La fórmula de F para el análisis de varianza es la siguiente:

$$F = MS_E / MS_D$$

La media de cuadrados, entre y dentro de los grupos, se obtiene dividiendo las sumas de cuadrados respectivas entre los grados de libertad entre y dentro de los grupos. El siguiente cuadro ilustra lo que se dice:

Fuente de variación	SS	gl	MS
Variación entre las combinaciones	306,26	3	102,9
Variación dentro de las combinaciones	740,86	316	2,34
Totales	1047,12	319	

Cuadro N° 12

SS es la abreviación de suma de cuadrados (del inglés *sume of square*). MS es la abreviación de media de cuadrados (del inglés *mean of square*); gl es la abreviación de grados de libertad.

Los grados de libertad se obtienen de la manera que a continuación se describe. Grados de libertad entre las combinaciones: como quiera que existen cuatro combinaciones o cuatro grupos de tratamiento, el investigador ha trabajado con una tabla de contingencia 2x2, por tanto los grados de libertad, entre las combinaciones, son 4 – 1 = 3.

Grados de libertad dentro de las combinaciones: como quiera que en cada uno de los tratamientos o combinaciones existen 80 alumnos, se resta 80 – 1 = 79, en cada combinación. Como existen cuatro combinaciones se tiene: 79 x 4 = 316 grados de libertad dentro de las combinaciones.

Las operaciones para obtener F son las siguientes:

$$F = MS_E / MS_D$$

$$F = 102,09 / 2,34 = 43,63$$

43,63 es el valor hallado de F. Este valor se compara con el valor tabulado de F. El valor tabulado de F aparece en el cuadro N° 13. En este caso, para 3 grados de libertad entre las combinaciones y 316 grados de libertad dentro de las combinaciones, el valor tabulado es 2,62. Entonces se observa que:

$$43,63 > 2,63$$

Como el valor hallado es mayor que el valor tabulado, se adopta la decisión de rechazar la hipótesis nula, que sostiene: *El eficiente desempeño docente,* (A_2), *y la aplicación de métodos didácticos centrados en el aprendizaje,* (B_2), *no se relacionan significativamente con el nivel de rendimiento académico de estudiantes del Ciclo Básico de Obstetricia* y, por el contrario, se acepta la hipótesis alterna que sostiene: *El eficiente desempeño docente,* (A_2), *y la aplicación de métodos didácticos centrados en el aprendizaje,*(B_2),*influyen significativamente en el nivel de rendimiento académico de estudiantes del Ciclo Básico de Obstetricia.*

DISTRIBUCIÓN DE F (95% de confianza)

dentro de las combinaciones	Entre las combinaciones					
Gl	1	2	3	4	5	6
1	161,40	199,50	215,70	224,60	230,20	234,00
2	18,51	19,00	19,16	19,25	19,30	19,33
3	10,13	9,55	9,28	9,12	9,01	8,94
4	7,71	6,94	6,59	6,39	6,26	6,16
5	6,61	5,79	5,41	5,19	5,05	4,95
6	5,99	5,14	4,76	4,53	4,39	4,28
7	5,59	4,74	4,35	4,12	3,97	3,87
8	5,32	4,46	4,07	3,84	3,69	3,58
9	5,12	4,26	3,86	3,63	3,48	3,37
10	4,96	4,10	3,71	3,48	3,33	3,22
11	4,84	3,98	3,59	3,36	3,20	3,09
12	4,75	3,89	3,49	3,26	3,11	3,00
13	4,67	3,81	3,41	3,18	3,03	2,92
14	4,60	3,74	3,34	3,11	2,96	2,85
15	4,54	3,68	3,29	3,06	2,90	2,79
16	4,49	3,63	3,24	3,01	2,85	2,74
17	4,45	3,59	3,20	2,96	2,81	2,70
18	4,41	3,55	3,16	2,93	2,77	2,66
18	4,38	3,52	3,13	2,90	2,74	2,63
20	4,35	3,49	3,10	2,87	2,71	2,60
21	4,32	3,47	3,07	2,84	2,68	2,57
22	4,30	3,44	3,05	2,82	2,66	2,55
oo	3,84	3,00	2,60	2,37	2,21	2,10

Cuadro N° 13

8. ANÁLISIS DE REGRESIÓN MÚLTIPLE

El Análisis de Regresión Múltiple es otra técnica para contrastar hipótesis con dos o más variables independientes. El diseño o la estrategia, en este caso, es factorial o multivariado. Se emplea el Análisis de Regresión Múltiple cuando las variables independientes se encuentran relacionadas entre sí. Esto significa que existe covariación entre las variables independientes. El siguiente ejemplo es una investigación realizada por el autor.

Estudio de la correlación entre el índice académico, la organización del tiempo libre y la afinidad de la actividad laboral con los estudios, con respecto al éxito académico de alumnos de post-grado de la Universidad Nacional Mayor de San Marcos.

El objetivo principal de este estudio es predecir el éxito académico. Esto es posible empleando el Análisis de Regresión Múltiple. Esta prueba, según Kerlinger, se basa en la misma idea de la correlación simple excepto cuando X, la variable independiente, es más de una. En el análisis de regresión simple, la variable X se emplea para predecir Y. En el análisis de regresión múltiple, se emplean las variables X_1, X_2, X_3, X_n para predecir Y. La organización de los datos y los cálculos se hacen de tal manera que se pueda obtener la predicción más acertada a partir de la combinación de las variables independientes. Es decir, en lugar de afirmar: si se tiene X, se obtendrá Y, se afirma: si se tiene X_1, X_2, X_3, X_n, se obtendrá Y. El resultado de los cálculos expresará la bondad o exactitud de la predicción y medirá la fracción de la varianza de Y que depende de la mejor combinación de las variables independientes.

Es oportuno aplicar el Análisis de Regresión Múltiple cuando las variables se presentan concurrente y concomitantemente, esto significa que cuando al presentarse juntas, se potencian los efectos de cada una de las variables: los efectos conjuntos de las variables son mayores de lo que podrían producir, por separado, cada una de ellas. Y como las variables independientes están relacionadas entre sí, ya no se pueden separar para estudiar los efectos de cada una de ellas, como se hace en el Análisis de Varianza, sino que se trata de hallar la mejor combinación posible de los efectos conjuntos de las variables independientes.

Para estudiar los datos empleando el Análisis de Regresión Múltiple, es necesario reconocer el principio de la partición de la varianza, es decir reconocer la medida en que *Y* (variable dependiente) se debe a cada una de las variables independientes. Pero es el caso que las variables independientes del presente estudio no solamente están correlacionadas cada una con la variable dependiente, sino que también se encuentran correlacionadas entre sí. Esta particularidad de las variables va a determinar el proceso a seguir.

Problema

¿Qué grado de correlación existe entre las variables índice académico, organización del tiempo libre y afinidad de la actividad laboral con los estudios, con respecto al éxito académico de estudiantes de Post Grado de la Universidad Nacional Mayor de San Marcos?

Operacionalización de las variables

Índice Académico: Es el promedio ponderado de las puntuaciones que alcanzaron estudiantes de post grado, cuando cursaron estudios de pre grado. Es variable continua que puede medirse en el nivel de intervalo. Ha sido medida en la escala vigesimal.

Organización del tiempo libre: Es la puntuación en la escala vigesimal que alcanzan los estudiantes de post grado a partir del análisis de componentes tales como: ocupaciones remuneradas, horas libres, horas de estudio dedicado a la lectura y posesión personal de libros. Es variable continua que puede medirse en el nivel de intervalo. Ha sido medida en la escala vigesimal.

Afinidad de la actividad laboral con los estudios: Se determina por el análisis de dos componentes: la declaración del propio sujeto y la interpretación de las tareas que cumple en su actividad laboral. Es variable continua que puede medirse en el nivel de intervalo. Ha sido medida en la escala vigesimal.

Éxito académico: Es el promedio ponderado de las calificaciones obtenidas por estudiantes de post grado de la UNMSM. Es variable continua que puede medirse en el nivel de intervalo. Ha sido medida en la escala vigesimal.

Hipótesis alterna

Existe un coeficiente de correlación, estadísticamente significativo, al nivel de 0,05, entre las variables índice académico, organización del tiempo libre y afinidad de la actividad laboral con los estudios, con respecto al éxito académico de estudiantes de post grado de la Universidad Nacional Mayor de San Marcos.

En términos formales, la hipótesis se puede expresar en términos de la siguiente función:

$$Y = f (X_1 X_2 X_3)$$

Hipótesis nula

El índice académico, la organización del tiempo libre y la afinidad de la actividad laboral con los estudios no están correlacionados con el éxito académico de estudiantes de post grado de la UNMSM.

Correlación simple

Como se ha señalado anteriormente, las variables independientes de esta investigación se encuentran correlacionadas entre sí. Esto ha sido demostrado luego de hallar los índices de correlación simple entre tales variables. Para ello se empleó la fórmula de correlación producto momento de Spearman, para encontrar los coeficientes de correlación entre:

X_1 X_2

X_1 X_3

X_2 X_3

antes de hallar la correlación entre:

YX_1

YX_2

YX_3

Covarianza

Elevando al cuadrado cada uno de los coeficientes de correlación simple se obtuvo la covarianza, es decir, los porcentajes de la varianza compartida entre las tres variables independientes y la dependiente. Los datos obtenidos son los siguientes:

	Y	X_1	X_2
X_1	0,16		
X_2	0,03	0,03	
X_3	0,08	0,08	0,11

Cuadro N° 14

La presencia de correlación entre las variables independientes significa que éstas se traslapan entre sí, no sólo en el plano bidimensional, sino en múltiples dimensiones. Este fenómeno complica el análisis de los datos y no se puede graficar en el plano bidimensional del papel. Sin embargo, en el Gráfico Nº 9, se trata se representar los traslapamientos entre cada una de las variables y la zona sombreada viene a ser el sector de la varianza de Y de la que es responsable la presencia conjunta, concurrente y concomitante de las variables independientes.

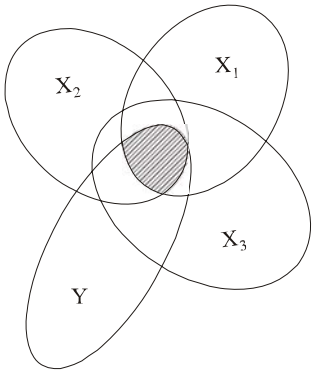


Gráfico N° 10

La varianza en este caso, no es sinónimo de la medida de las diferencias entre puntuaciones o rango de las puntuaciones entre individuos o grupos, sino se entiende como la variabilidad de los valores de una variable dependiente que se debe o se explica por la presencia de otras variables, llamadas independientes.

Si las variables independientes no estuviesen correlacionadas entre sí, sería suficiente sumar los porcentajes de la varianza que explican, por separado, cada una de las variables independientes. De esta manera se afirmaría que el 49% de la varianza de Y es explicada por X_1 X_2 y X_3. Pero éste no es el caso.

Predicción

El análisis de regresión múltiple es un método estadístico que permite interpretar el fenómeno descrito y conduce a la predicción de los valores de la variable dependiente. La predicción es mejor en la medida en que la correlación entre las variables independientes entre sí sea baja, como en el presente caso: correlación entre X_1 y X_2 = 0,3; correlación entre X_1 y X_3 = 0,8, y correlación entre X_2 y X_3 = 0,11.

Mediante programas de computadora, se resolvieron las siguientes ecuaciones del modelo estadístico de regresión múltiple para tres variables independientes y una dependiente.

$$\Sigma Y = a_0 n + a_1 \Sigma X_1 + a_2 \Sigma X_2 + a_3 \Sigma X_3$$

$$\Sigma X_1 Y = a_0 \Sigma X_1 + a_1 \Sigma X_1^2 + a_2 \Sigma X_1 X_2 + a_3 \Sigma X_1 X_3$$

$\Sigma X_2 Y = a_0 \Sigma X_2 + a_1 \Sigma X_1 X_2 + a_2 \Sigma X_2^2 + a_3 \Sigma X_2 X_3$

$\Sigma X_3 Y = a_0 \Sigma X_3 + a_1 \Sigma X_1 X_3 + a_2 \Sigma X_2 X_3 + a_3 \Sigma X_3^2$

Los valores calculados para resolver las ecuaciones fueron los siguientes:

$n = 145$
$\Sigma X_1 = 2023,57$
$\Sigma X_2 = 1663$
$\Sigma X_3 = 1900$
$\Sigma Y = 2094,71$
$\Sigma X_1^2 = 28393,28$
$\Sigma X_2^2 = 19939$
$\Sigma X_3^2 = 25184$
$\Sigma Y^2 = 30410,96$
$\Sigma X_1 X_2 = 23266,34$
$\Sigma X_1 X_3 = 36570,90$
$\Sigma X_2 X_3 = 21959$

a_0 = constante
a_1 = coeficiente de regresión de X_1
a_2 = coeficiente de regresión de X_2
a_3 = coeficiente de regresión de X_3

Resueltas las ecuaciones del modelo estadístico, se pudo resolver la siguiente ecuación de predicción:

$$Y = a_0 + X_1 a_1 + X_2 a_2 + X_3 a_3$$

En donde

a_0 : Constante = 7,7844
a_1 : Coeficiente de regresión de X_1 = 0,3421
a_2 : Coeficiente de regresión de X_2 = 0,0282
a_3 : Coeficiente de regresión de X_3 = 0,1194

Conociendo el valor de la constante y de los coeficientes de regresión de las variables independientes, es posible predecir el valor de Y, es decir, la puntuación que obtendría cualquier estudiante al término de sus estudios de post grado, siempre y cuando se tengan de él sus puntajes alcanzados en las variables X_1, X_2 y X_3. Así, si se seleccionara aleatoriamente a cualquier sujeto de los 145 de la muestra, por ejemplo al Nº 40, se podría predecir su éxito académico utilizando la ecuación de predicción.

$\hat{Y}$ del sujeto 40 = 7,7844 + (13,87 x 0,3421) + (13 x 0,0282) + (14 x 0,1194)

$\hat{Y}$ del sujeto 40 = 7,7844 + 4,74927 + 0,3666 + 1,6716

$\hat{Y}$ del sujeto 40 = 14,57
Y del sujeto 40 = 14,19

Discrepancia entre $\hat{Y}$ y Y = –0,38
Otro caso:

$\hat{Y}$ del sujeto 101 = 7,7844 + (15,03 x 0,3421) + (12 x 0,0282) + (14 x 0,1194)

$\hat{Y}$ del sujeto 101 = 7,7844 + 5,141673 + 0,3384 + 1,6716

$\hat{Y}$ del sujeto 101 = 14,94
Y del sujeto 101 = 14,50

Discrepancia entre $\hat{Y}$ y Y = 0,44

Coeficiente de correlación múltiple

Si se obtuvieran todas las puntuaciones predichas ($\hat{Y}$) empleando el procedimiento descrito, sería posible correlacionarlas con las puntuaciones reales (Y). Este coeficiente de correlación no es el mismo que el normal, *r*, que se obtiene correlacionando dos series de puntuaciones reales, sino se denomina Coeficiente de Correlación Múltiple y su símbolo es *R* y se obtiene correlacionando las puntuaciones reales de Y con las puntuaciones predichas de Y y expresa la magnitud de la relación entre la mejor combinación posible de todas las variables independientes con respecto a la variable dependiente. En el presente caso, el valor de R es:

R = 0,4427

Mediante procesos de computadora se calcularon los valores predichos de Y de todos los 145 sujetos de la muestra, por tanto, el valor de R se ha sido hallado luego de correlacionar la serie de puntuaciones predichas de Y con las puntuaciones reales de Y.

Este es el dato más importante de la investigación, porque a partir de él se va a adoptar la decisión con respecto a la hipótesis nula.

El coeficiente de regresión múltiple, R, elevado al cuadrado, indica la varianza compartida entre $\hat{Y}$ y Y, es decir R^2 es el porcentaje de la varianza de Y de la cual es responsable la concurrencia de las tres variables independientes.

En este caso, dicho valor es:

$$R^2 = 0{,}4427^2 = 0{,}1956$$

Redondeando sería = 0,20

Y en términos porcentuales sería = 20%

Que significa que las tres variables independientes: índice académico, organización del tiempo libre y afinidad de la actividad laboral con los estudios, son responsables del 20% de la varianza de Y: éxito académico en los estudios de post grado. Como es natural, otros factores, no estudiados en esta investigación, serían los responsables del 80% de la varianza restante. Lo dicho se puede apreciar más claramente en el siguiente gráfico:

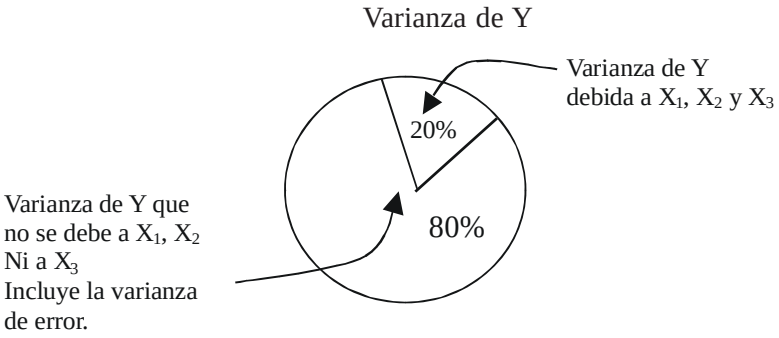


Gráfico N° 10

Grados de libertad

En este estudio, como se ha trabajado con una muestra de 145 estudiantes y con cuatro variables, los grados de libertad se establecen del siguiente modo:

$$K - 4 = 145 - 4 = 141$$

$$L - 1 = 4 - 1 = 3$$

Prueba F

Para el contraste de la hipótesis, luego de realizar el análisis de regresión múltiple, se empleó la Prueba F, al nivel de significación de 0,05.

Cuando se hace el análisis de regresión múltiple, la fórmula de la prueba F es la siguiente:

$$F = \frac{R^2 / K}{\left(1 - R^2\right) / \left(N - K - 1\right)}$$

En donde:

R^2 = Coeficiente de regresión múltiple elevado al cuadrado.

K = Número de variables independientes.

N = Tamaño de la muestra.

Al hacer las operaciones se obtiene:

$$F = \frac{0,4427^2 / 3}{\left(1 - 0,4427^2\right) / (145 - 3 - 1)}$$

$$F = \frac{0,0653}{0,0057021} = 11,4519$$

El valor encontrado, luego de reemplazar los términos de la fórmula es 11,4519, superior a 2,60, valor crítico para 141 grados de libertad y tres variables independientes, tal como puede verse en el Cuadro N° 13, Distribución de F, con 95% de confianza. En consecuencia se rechaza la hipótesis nula al nivel de 0,05 que sostiene: *El índice académico, la organización del tiempo libre y la afinidad de la actividad laboral con los estudios no están correlacionados con el éxito académico de estudiantes de post grado de la UNMSM.*

En cambio, se acepta la hipótesis alterna que sostiene: *Existe un coeficiente de correlación, estadísticamente significativo, al nivel de 0,05, entre las variables índice académico, organización del tiempo libre y afinidad de la actividad laboral con los estudios, con respecto al éxito académico de estudiantes de post grado de la Universidad Nacional Mayor de San Marcos.*

9. PRUEBA t DE STUDENT

La Prueba t de Student, llamada también prueba de las diferencias de medias, es otra prueba muy usada para contrastar hipótesis. Cuando los investigadores adoptan estrategias de prueba de hipótesis como consecuencia de las cuales se obtienen dos promedios, en cada uno de los grupos, ya sean éstos experimental o de control, lo que deben es determinar si la diferencia entre tales promedios hallados se debe a hechos fortuitos o si tal diferencia se ha producido como efecto de la influencia de la variable independiente que se está estudiando. En este caso, el investigador, con los datos que obtiene, realiza las operaciones respectivas.

La limitación de esta prueba es que no funciona con muestras grandes, sólo con muestras pequeñas. Algunos autores sostienen que el número de la muestra no debe pasar de 20.

A continuación se presenta el estudio realizado por Marco Sanabria Montañez, titulado: *Influencia del Seminario y la Clase Magistral en el Rendimiento Académico de alumnos de la E. A. P. de Economía de la U. N. M. S. M.*

Problema:

¿Existen diferencias significativas en el éxito académico alcanzado por alumnos de la asignatura de Economía General I de la Escuela Académico Profesional de Economía de la Universidad Nacional Mayor de San Marcos que realizaron sus estudios con el método de Seminario, con respecto al grupo de alumnos que cursaron con el método de la Clase Magistral?

Definición de términos básicos

Clase Magistral: Es un tiempo de enseñanza ocupado entera o principalmente por la exposición continua de un conferenciante. Los estudiantes pueden tener la oportunidad de preguntar o participar en una pequeña discusión pero, por lo general, no hacen otra cosa que escuchar y tomar nota.

Seminario: Es una técnica didáctica en el que se reúnen un número pequeño de miembros que estudian e investigan sistemáticamente un tema con el objeto de lograr el conocimiento completo y específico de una materia.

Índice Académico: Es el resultado cuantitativo en una escala de valores, que determina el nivel alcanzado en el rendimiento académico por un estudiante.

Hipótesis general

El éxito académico de los alumnos que desarrollan la asignatura de Economía General I con el método de Seminario, es significativamente diferente al Índice Académico de los alumnos que estudian la asignatura de Economía General I con el método de Clase Magistral en la E. A. P. de Economía de la UNMSM

Hipótesis específica 1

El éxito académico de los alumnos de la E. A. P. de Economía en la UNMSM que desarrollan la asignatura de economía se incrementa significativamente con la aplicación del método didáctico de Seminario.

Hipótesis específica 2

El éxito académico de los alumnos de la E. A. P. de Economía en la UNMSM que desarrollan la asignatura de economía, no se incrementa significativamente con la aplicación del método didáctico de Clase Magistral.

Variables e indicadores

Las variables consideradas para la presente investigación son las siguientes:

Variable independiente: Método didáctico

Variación:

a) Método didáctico de Clase Magistral (Z),

b) Método didáctico de Seminario (X)

Variable dependiente: éxito Académico (O)

Variación:

a) Éxito académico alto.

b) Éxito académico bajo.

Operacionalización de la variable índice académico

Éxito académico Alto: Cuando el promedio obtenido por la muestra es 12,31 ó más.

Éxito académico Bajo: Cuando el promedio de la muestra es menor que 12,31.

Estos valores se han calculado mediante el método estadístico de estimación puntual de la diferencia de dos medias, en donde el estimador es $(\bar{O}_2 - \bar{O}_4)$.

La desviación estándar del estimador puntual de la diferencia de medias se obtiene mediante la aplicación de la siguiente fórmula:

$$\sigma_{(\bar{O}_2-\bar{O}_4)} = \sqrt{\frac{\sigma^2_{\bar{O}_2}}{n_{\bar{O}_2}} + \frac{\sigma^2_{\bar{O}_4}}{n_{\bar{O}_4}}}$$

Para el cálculo de la Cota de Error se ha utilizado la fórmula:

$$2\left(\sigma_{(\bar{O}_2-\bar{O}_4)}\right)$$

Con las fórmulas mencionadas se han realizado las operaciones correspondientes y se tiene:

$$(\bar{O}_2 - \bar{O}_4) = 12{,}8 - 8{,}62 = 4{,}18$$

$$\sigma(\bar{O}_2 - \bar{O}_4) = 0{,}7992$$

Cota de error = 2 x 0,7992 = 1,5981

El promedio obtenido de ambas muestras es: 10,71, por la operación de la semisuma de los promedios de las muestras del Seminario y Clase Magistral.

A este promedio se adiciona la cota de error, tendremos 10,71 + 1,5981 entonces se obtiene el valor crítico 12,3081 que es el límite de los éxitos académicos.

Estrategia para la prueba de hipótesis

La presente es una investigación con medición previa y posterior con grupo de control. Las unidades de análisis, tanto en el grupo experimental como en el grupo de control, se encuentran asignadas en grupos ya establecidos. El esquema de diseño es cuasi experimental, y se expresa de la siguiente manera:

Grupo que desarrolla Seminario:	O_1	X	O_2
Grupo que desarrolla Clase Magistral:	O_3	Z	O_4

En donde:

X: Método de Seminario.

Z: Método de Clase Magistral.

O_1: Medición del éxito Académico del grupo que desarrolla el Seminario antes del experimento.

$O_{2:}$ Medición del éxito Académico del grupo que desarrolla el Seminario después del experimento.

O_3: Medición del éxito Académico del grupo de que desarrolla la Clase Magistral antes del experimento.

O_4: Medición del éxito Académico del grupo que desarrolla la Clase Magistral después del experimento.

Análisis del Pre-Test

El empleo del pre test ha tenido como finalidad determinar el nivel de conocimientos con el que inician el experimento el grupo al que se aplica el seminario (O_1), y el grupo al que se aplica la Clase Magistral (O_3).

Para el del cálculo del promedio se utilizó la siguiente fórmula:

$$\overline{O} = \frac{\sum_{i=1}^{n} 0\, f_i}{n}$$

Para el cálculo de la desviación estándar se utilizó la formula:

$$S_{\overline{0_1}} = \sqrt{\frac{\sum f_1\ 0_{1_i}^2 - \frac{\left(\sum f_1\, o_{1_i}\right)^2}{n}}{n-1}}$$

Hechas las operaciones, los datos hallados son los siguientes:

Pre Test Estadígrafos	O_1	O_3
Promedio	3,800	4,4100
Desviación Estándar	1,1464	1,1657
Total Participantes	15,000	37,0000

Cuadro N° 15

Estimador puntual de la diferencia de medias

Permite identificar un punto crítico al que se le conoce con el nombre de Cota de Error. La decisión se toma en el sentido de que si la Cota de Error es mayor que la diferencia de medias no hay diferencia significativa en los promedios analizados y viceversa.

Cálculo del Estimador $\left(\overline{O}_1 - \overline{O}_3\right)$ = 0,61

Desviación estándar de la diferencia de medias obtenidas es

$$\sigma\ \left(\overline{O}_1 - \overline{O}_3\right) = 0{,}352628$$

La cota de error = 0,705

La cota de error: 0,705 es mayor que la diferencia de las medias: 0,61, por lo tanto los promedios obtenidos por los alumnos en el pre test no difieren significativamente.

Análisis Post Test

Pre Test Estadígrafos	O_1	O_3
Promedio	3,800	4,4100
Desviación Estándar	1,1464	1,1657
Total Participantes	15,000	37,0000

Cuadro N° 16

Contraste de Hipótesis

Mason, Lind y Marchal mencionan que "*existe un procedimiento de cinco pasos que sistematiza la prueba de hipótesis. Al llegar al paso 5, se tiene ya la capacidad de tomar la decisión de rechazar o no la hipótesis*"[1]. Atendiendo este planteamiento que a criterio del autor de esta investigación es el más coherente, se desarrolló el proceso.

Paso 1: Plantear la Hipótesis Nula (H_O) y la Hipótesis Alterna (H_1):

Hipótesis General

Hipótesis Nula (H_0)

El éxito Académico de los alumnos que desarrollan la asignatura de Economía General I con el método de Seminario (O_2), *no es significativamente diferente al éxito Académico de los alumnos que estudian la asignatura de Economía General I con el método de Clase Magistral* (O_4) *en la E. A. P. de Economía de la UNMSM.*

Lo dicho, formalmente se expresa mediante la siguiente ecuación:

$$H_0: \ O_2 = O_4$$

Hipótesis Alterna (H_1)

El éxito Académico de los alumnos que desarrollan la asignatura de Economía General I con el método de Seminario (O_2), es significativamente diferente del éxito Académico de los alumnos que estudian la asignatura de Economía General I con el método de Clase Magistral (O_4) en la E. A. P. de Economía de la UNMSM.

Lo dicho, formalmente se expresa mediante la siguiente inecuación:

$$H_1: O_2 > O_4$$

Paso 2: Seleccionar el nivel de significación:

$$\alpha = 0.05$$

Paso 3: Escoger el valor estadístico de prueba:

$$Z = \frac{\overline{O}_2 - \overline{O}_4}{\sqrt{\frac{S_2^2}{n_2} + \frac{S_4^2}{n_4}}}$$

Paso 4: Formular la regla de decisión:

Para la Hipótesis General:

Rechazar la Hipótesis nula si: $Z > 1,96$

Paso 5: Adoptar una decisión:

Post test / Estadígrafos	O_2	O_4
Promedio	12,8000	8,620
Desviación Estándar	2,1112	3,554
Total Participantes	15,0000	37,00

Cuadro N° 17

$$Z = \frac{12.8 - 8.62}{\sqrt{\frac{4.4571}{15} + \frac{12.6306}{37}}}$$

$Z = 5,23$ (Z calculada)

Se rechaza la hipótesis nula y se acepta la hipótesis alterna que es la siguiente:

El éxito Académico de los alumnos que desarrollan la asignatura de Economía General I con el método de Seminario (O_2), *es significativamente diferente a éxito Académico de los alumnos que estudian la asignatura de Economía General I con el método de Clase Magistral* (O_4) *en la E. A. P. de Economía de la UNMSM.*

Se rechaza la hipótesis nula de la Hipótesis específica y se acepta la hipótesis:

El éxito académico de los alumnos de la E. A. P. de Economía en la UNMSM que desarrollan la asignatura de economía se incrementa significativamente con la aplicación del método didáctico de Seminario.

Conclusiones

En la presente investigación se ha llegado a las siguientes conclusiones:

1. Existe evidencia empírica que el éxito académico que se obtiene aplicando el método didáctico de Seminario difiere significativamente al éxito académico obtenido con la aplicación del método didáctico de Clase Magistral.
2. Los estudiantes a los que se les aplicó el método didáctico de seminario alcanzaron un éxito académico alto, mientras que los estudiantes a los que se les aplicó el método didáctico de Clase Magistral alcanzaron un éxito académico bajo.

DISTRIBUCIÓN DE t

gl	0,20	0,10	0,05	0,02	0,01	0,002
1	3,08	6,71	12,70	31,80	63,60	318,30
2	1,89	2,92	4,30	6,97	9,93	22,30
3	1,64	2,35	3,18	4,54	5,84	10,20
4	1,53	2,13	2,78	3,75	4,60	7,17
5	1,48	2,02	2,57	3,37	4,03	5,89
6	1,44	1,94	2,45	3,14	3,71	5,21
7	1,42	1,90	2,37	3,00	3,50	4,79
8	1,40	1,86	2,31	2,90	3,36	4,50
9	1,38	1,83	2,26	2,82	3,25	4,30
10	1,37	1,81	2,23	2,76	3,17	4,14
11	1,36	1,80	2,20	2,72	3,11	4,03
12	1,36	1,78	2,18	2,68	3,06	3,93
13	1,35	1,77	2,16	2,65	3,01	3,85
14	1,35	1,76	2,15	2,62	2,98	3,79
15	1,34	1,75	2,13	2,60	2,95	3,73
16	1,34	1,75	2,12	2,58	2,92	3,69
17	1,33	1,74	2,11	2,57	2,90	3,65
18	1,33	1,73	2,10	2,55	2,88	3,61
19	1,33	1,73	2,09	2,54	2,86	3,58
20	1,33	1,73	2,09	2,53	2,85	3,55
21	1,32	1,72	2,08	2,52	2,83	3,53
22	1,32	1,72	2,07	2,51	2,82	3,51
oo	1,28	1,64	1,96	2,33	2,58	3,09

Cuadro N° 18

LECTURA COMPLEMENTARIA

PRUEBA DE HIPÓTESIS Y ERROR ESTÁNDAR

Fred Kerlinger*

El error estándar, como estimación de las fluctuaciones del azar es la medida en contra de la cual los resultados de los experimentos son probados. ¿Hay alguna diferencia entre las medias de dos grupos experimentales? Si esto sucede, ¿Es la diferencia una diferencia "real", o es sólo una consecuencia de las muchas diferencias relativamente pequeñas que pudieron haber surgido por el azar?

Para contestar estas preguntas, se calcula el error estándar de las diferencias entre las medias y la diferencia obtenida se compara con el error estándar. Si es suficientemente más grande que el error estándar, se dice que es una diferencia "significativa".

Un razonamiento similar se puede aplicar a cualquier estadístico. Por lo tanto, hay muchos errores estándar: de coeficientes de correlación, de las diferencias entre las medias, de las medias, de las medianas, de las proporciones, etcétera. El propósito en este capítulo es examinar la idea general del error estándar y ver cómo se prueban las hipótesis usando el error estándar.

EJEMPLO: DIFERENCIAS ENTRE MEDIAS

Un problema particularmente difícil en la psicología contemporánea se centra en contestar la pregunta de si la conducta es controlada por factores situacionales o ambientales, o por disposiciones del individuo. McGee y Snyder, usando una presunta diferencia entre las personas que le ponían sal a sus alimentos antes de probarlos y aquellas que probaban sus alimentos antes de agregarles sal, elaboraron la hipótesis según la cual las personas que interpretan su conducta de forma ordenada, le agregaban sal a sus alimentos antes de probarlos, mientras que aquellas personas que interpretaban su conducta en forma situacional probaban sus alimentos antes de agregarles sal.

---

* Tomado de: KERLINGER, Fred N.: *Investigación del comportamiento.* México, D.F. McGraw-Hill, 1988. pp. 210-226

Además, señalaron que las primeras personas adjudicarían más rasgos a sí mismos que las otras personas. Encontraron que los primeros, los "que le agregaban sal", se atribuían una media de 14,87 rasgos a sí mismos mientras que el segundo grupo, los "que prueban antes los alimentos", se atribuían una media de 6,90 rasgos a sí mismos. La dirección de la diferencia fue como había sido esperada por los autores. ¿Es el tamaño de la diferencia entre las medias, 7,97, lo suficientemente grande para apoyar el reclamo de los autores de que su hipótesis fue apoyada? Una prueba de significancia estadística de esta diferencia muestra que ésta fue altamente significativa.

La idea del presente ejemplo en este contexto es que la diferencia entre las medias fue evaluada para determinar la significancia estadística, con un error estándar. El error estándar en este caso fue el error estándar de las diferencias entre las dos medias. Se observó que esta diferencia era significativa. Esto quiere decir que las personas que percibieron la conducta como influida por los rasgos individuales, tendieron a agregarle sal a sus alimentos antes de probarlos, mientras que aquéllas cuya percepción está más orientada de manera ambiental, probaron sus alimentos antes de agregarles sal. (Esta afirmación es una generalización de la original.) Ahora, examine un ejemplo en el cual la diferencia entre las medias no fue significativa.

Gates y Taylor, en un estudio inicial muy conocido sobre la transmisión de entrenamiento, organizaron dos grupos de 16 alumnos cada uno. Al grupo experimental se le proporcionó una práctica en memoria de dígitos; al control no se le proporcionó dicho entrenamiento. La mejoría del promedio del grupo experimental inmediatamente después del periodo de entrenamiento, fue de 2,00; la ganancia media del grupo control fue de ,67, una diferencia media de 1,33. Cuatro a cinco meses más tarde, los niños de ambos grupos fueron evaluados de nuevo. La media de las calificaciones del grupo experimental fue de 4,71; la media del grupo control fue sorprendentemente de 4,77. El incremento de las medias sobre las pruebas iniciales fue de ,35 y ,36. Las pruebas estadísticas son innecesarias con datos como estos.

DIFERENCIAS ABSOLUTAS Y RELATIVAS

Como las diferencias entre los estadísticos, especialmente entre las medias, son evaluadas y reportadas en gran parte en la literatura, se deben analizar con detalle los tamaños absoluto y relativo de tales estadísticos. Aunque en la discusión se usan como ejemplos diferencias entre medias, las mismas ideas se aplican a diferencias entre proporciones, coeficientes de correlación, etcétera.

En un estudio hecho por Goldberg, Gottesdiner y Abramson, a las mujeres que fueron consideradas como simpatizantes del movimiento de liberación femenina se

les calificó como menos atractivas que las consideradas como no simpatizantes del movimiento. Las no simpatizantes tuvieron una media de 2,86 y las simpatizantes 2,75, una diferencia de ,11, la cual fue estadísticamente significativa. ¿Es significativa una diferencia tan pequeña? Compárese esta pequeña diferencia con aquélla de medias entre los grupos experimental y control obtenida por Mann y Janis en su estudio de los efectos a largo plazo de la simulación sobre el fumar: 13,50 y 5,20. (Éstas son medias de los decrementos en los números de cigarrillos fumados cada día; la diferencia es estadísticamente significativa).

En realidad aquí hay dos problemas: uno de ellos es del tamaño absoluto y relativo de la diferencia y otro de la significancia práctica o "real" en contraste con la significancia estadística. La que en apariencia es una diferencia muy pequeña puede, bajo un examen minucioso, no serlo tanto. En el estudio de Goldberg y cols., para estar seguros, la diferencia de ,11 es a lo mejor trivial, aunque estadísticamente significativa. El ,11 se derivó de una escala de atracción de cinco puntos y es, por lo tanto, en realidad pequeña. Ahora tómese un ejemplo completamente diferente, de un estudio de Miller y Dicara sobre el condicionamiento instrumental de la secreción urinaria. Las medias de un grupo de ratas antes y después del entrenamiento en la secreción de la orina fueron ,017 y ,028 y la diferencia fue muy estadísticamente significativa.

Sin embargo la diferencia fue de sólo ,011. ¿No es esta diferencia tan pequeña como para prestarle atención? Sin embargo, la naturaleza de las medidas tiene que tomarse en cuenta. Las pequeñas medias de ,017 y ,028 se obtuvieron de medidas de secreción de las ratas. Cuando se toma en cuenta el tamaño de la vejiga de las ratas y que la diferencia de medias de ,011 fue producida a través del condicionamiento instrumental (recompensa por secreción urinaria), el significado de la diferencia es dramática: ¡es inclusive muy grande!

No se debería ser demasiado entusiasta por las diferencias de medias como ,20, ,15, ,08, etc., sino que se ha de ser inteligentes respecto a ellas. Supóngase que una pequeña diferencia se informa como estadísticamente significativa y uno piensa que es algo ridícula. Sin embargo, supóngase que dicha diferencia fue la de medias entre los pesos de la corteza cerebral de un grupo de ratas criadas en medios ricos y en medios pobres en los primeros días de vida. El obtener *cualquier* diferencia en el peso del cerebro debido al tipo de ambiente es un logro extraordinario y, desde luego, un descubrimiento científico importante.

COEFICIENTES DE CORRELACIÓN

Los coeficientes de correlación abundan en las publicaciones científicas. Por lo tanto, se deben plantear preguntas acerca de la significancia de los coeficientes –y

de la "realidad" de las relaciones que estos expresan. Por ejemplo, para que un coeficiente de correlación calculado entre 30 pares de medidas sea estadísticamente significativo, tiene que andar alrededor de ,31 al nivel de significancia de ,05 y de ,42 al nivel de ,01. Con 100 pares de medidas el problema es menos agudo (la ley de los grandes números entra en juego). En este caso, para satisfacer el nivel ,05, es suficiente una r de ,16; para satisfacer el nivel ,01, basta una *r* de ,23. Si las *r* son menores a estos valores se considera que son no significativas.

Si se seleccionan, por ejemplo, 30 pares de números de una tabla de números aleatorios y se calcula la correlación entre ellos, en la teoría debería de estar cerca del cero. Resulta claro que debería de haber relaciones cercanas al cero entre los conjuntos de los números aleatorios, pero ocasionalmente los conjuntos de pares pueden producir *r* estadísticamente significativas y razonablemente altas debidas *al azar.* De cualquier forma, los coeficientes de correlación, así como las medias y las diferencias, tienen que ser ponderados para valorar la significancia estadística mediante su comparación con los errores estándares. Por fortuna, esto es fácil de hacer, ya que las r para cada nivel de significancia y para tamaños diferentes de la muestra se proporcionan en tablas en la mayor parte de los textos sobre estadística. Por lo tanto, con las r no es necesario calcular y usar el error estándar de unir. Sin embargo, el razonamiento que subyace a las tablas ha de ser entendido.

De los miles de coeficientes de correlación publicados en la literatura sobre investigación, una gran parte es de baja magnitud. ¿Qué tan bajo es bajo? ¿En qué punto es un coeficiente de correlación lo suficientemente bajo para no tratarlo con seriedad? Usualmente, las r menores a ,10 no se deben tomar muy en serio: una r de ,10 significa que ¡sólo un 1% (,10) de la varianza de *y* es compartida con *x*! Si una r de ,30, por otro lado, es estadísticamente significativa, puede ser importante porque apunta hacia una relación importante. Las entre ,20 y ,30 hacen el problema un poco más difícil. (Recuérdese que con N grandes, las r entre ,20 y ,30 son estadísticamente significativas.) Para estar seguros, una r de, por ejemplo, ,20 quiere decir que dos variables comparten sólo el 4% de sus varianzas, pero una r de 26 –comparten un 7% de la varianza– o incluso una de ,20 puede ser importante porque puede proporcionar un indicio muy valioso para una teoría e investigación subsecuente.

El problema es complejo. En la investigación básica, las correlaciones bajas (desde luego que éstas deben ser estadísticamente significativas) pueden enriquecer la teoría y la investigación. En la investigación aplicada, donde las predicciones son importantes, han aumentado los juicios de valor acerca de las correlaciones bajas y las cantidades triviales de la varianza compartida. Sin embargo, en la investigación básica el panorama es más complicado. Una conclusión es razonablemente justa: los coeficientes de correlación, como otros estadísticos, deben ser evaluados para determinar su significancia estadística.

PRUEBA DE HIPÓTESIS: HIPÓTESIS SUSTANTIVA Y NULA

El propósito principal en la investigación de la estadística inferencial es probar las hipótesis de investigación a través de la evaluación de las hipótesis estadísticas. Hablando en términos amplios, el científico usa dos tipos de hipótesis: sustantivas y estadísticas. Una hipótesis sustantiva es el tipo usual de hipótesis que se comentó en el capítulo 2, en la cual es expresada, una afirmación de la relación entre dos o más variables, por ejemplo, "entre mayor es la cohesión de un grupo, mayor es su influencia sobre sus miembros" es una hipótesis sustantiva.

La teoría de un investigador determina que esta variable está relacionada con la otra. La afirmación de esta relación es una hipótesis sustantiva. Estrictamente hablando, una hipótesis sustantiva no es susceptible de ser probada o evaluada. Tiene, primero, que ser traducida a términos operacionales. Una forma muy útil de probarla es mediante una *hipótesis estadística,* la cual es una afirmación, en términos estadísticos, de las relaciones estadísticas deducidas de las relaciones planteadas en la hipótesis sustantiva. Este es un planteamiento demasiado vago y necesita ser explicado. Una hipótesis estadística expresa un aspecto de la hipótesis sustantiva original en términos estadísticos y cuantitativos. $M_A > MB$, la media de A es mayor que la media de B; $r > +,20$, el coeficiente de correlación es mayor que ,20; $M_A > M_B > M_C$, al nivel de ,01; x^2 es significativa al nivel de ,05; etcétera.

Una hipótesis estadística es una predicción de los resultados de los procedimientos estadísticos usados para analizar los datos cuantitativos de un problema de investigación.

Sin embargo, las hipótesis estadísticas deben ser comparadas contra algo. No es posible evaluar simplemente una hipótesis estadística como está planteada. Esto es, no se evalúa directamente la proposición estadística de $M_A > M_B$ contra sí misma. Se evalúa comparándola con una proposición alternativa. Naturalmente, puede haber varias alternativas de $M_A > M_B$.

La alternativa usualmente seleccionada es la hipótesis nula, la cual fue inventada por Sir Ronald Fisher. La hipótesis nula es una proposición estadística en que se plantea, principalmente, que no hay relación entre las variables (del problema). La hipótesis *nula* dice, "estás en un error, no hay relación; deséchame si puedes". Esto lo dice en términos estadísticos como $M_A = M_B$ o $M_A - M_B = 0$; $r_{xy} = 0$; la x^2 no es significativa; la t no es significativa; etcétera.

Fisher dice, "puede decirse que cada experimento es sólo para darle la oportunidad a los hechos de rechazar la hipótesis nula". Dicho eso en una forma más adecuada, ¿qué es lo que quiere decir? Supóngase que se planteó una hipótesis acerca de los efectos de algunos métodos y se dice que el método A es superior al

método B. Si uno resuelve en una forma satisfactoria los problemas de definir lo que se quiere decir con "superior", de diseñar un experimento, etc., ahora debe especificar una hipótesis estadística. En este caso, se podría decir $M_A > M_B$ (la media del método A es, o será mayor que la media del método B en tal y tal medida de criterio). Supóngase que después del experimento las dos medias son 68 y 61, respectivamente. Esto parecería ser que la hipótesis sustantiva es apoyada ya que 68 > 61, o M_A es mayor que M_B . Sin embargo, como ya se sabe, ésto no es suficiente ya que esta diferencia puede ser una de las muchas diferencias posibles debidas al azar.

De hecho, se diseñó una hipótesis que puede ser llamada la hipótesis del azar: $M_A = M_B$, o $M_A – M_B = O$. Estas son hipótesis nulas. Lo que se hace, entonces, es escribir hipótesis. Primero se escriben hipótesis estadísticas que reflejen el significado operacional–experimental de la hipótesis sustantiva. Después se escribe la hipótesis nula en contra de la cual se compara el primer tipo de hipótesis. Aquí se presentan dos tipos de hipótesis adecuadamente etiquetadas:

$$H_1: \ M_A > M_B$$

$$H_0; \quad M_A = M_B$$

H_1 significa "Hipótesis 1". Hay a menudo más de una de esas hipótesis. Ellas son etiquetadas como H_1. H_2, H_3, etcétera. H_o significa "Hipótesis nula". Nótese que en este caso la hipótesis nula podría haber sido escrita como:

$$H_0: \ M_A – M_B = 0$$

Esta forma muestra de donde la hipótesis nula obtuvo su nombre: la diferencia entre M_A y M_B es igual a cero, pero no es muy manejable en esta forma especialmente cuando se prueban tres o más medias u otros estadísticos. $M_A = M_B$ es general y desde luego significa lo mismo que $M_A – M_B = 0$ y $M_B – M_A = O$. Nótese que se puede escribir fácilmente $M_A = M_B = M_C = ... = M_N$.

NATURALEZA GENERAL DE UN ERROR ESTÁNDAR

Si éste fuera el mejor de todos los posibles mundos de investigación, no habría error aleatorio. Y si no hubiera error aleatorio, no habría necesidad de pruebas de significancia estadística. De hecho, la palabra "significancia" carecería de sentido. Cualquier tipo de diferencia sería una diferencia "real". Lamentablemente este nunca es el caso. *Siempre* existen errores al azar (y errores sesgados, también), y en la investigación del comportamiento a menudo contribuyen sustancialmente a la varianza total. Los errores estándar son mediciones de este error y se utilizan, como ya se ha mencionado varias veces, a manera de una especie de patrón contra el cual se contrasta la varianza experimental o "variable".

El *error estándar* es la desviación estándar de la distribución muestral de cualquier medición dada –la media o el coeficiente de correlación, por ejemplo–. En la mayor parte de los casos no es posible conocer los valores del universo o población (parámetros); deben ser estimados a partir de mediciones muestrales, usualmente de muestras sencillas.

Supóngase que se extrae una muestra aleatoria de 100 niños del octavo grado de un sistema escolarizado determinado. Es difícil o imposible, por decir algo, medir todo el universo de los niños de octavo grado. Se calculan la media y la desviación estándar a partir de un examen aplicado a los niños y se encuentra que estos estadísticos son $M = 110$ *y* $SD = 10$. Una pregunta importante que debe plantearse es "¿Qué tan exacta es esta media?" O, si fuera a extraerse un gran número de muestras aleatorias de 100 alumnos de octavo grado con base en esta misma población, ¿las medias de estas muestras serán 110 ó alrededor de 110? Y, si caen alrededor de 110, ¿con qué aproximación? Lo que se hace, en efecto, es establecer una *distribución hipótetica de medias de las muestras*, todas calculadas a partir de muestras de 100 alumnos, cada una de las cuales se ha extraído de la población original de alumnos del octavo grado. Si pudiera calcularse la media de esta población *de medias*, o si se conociera cuál es, todo sería más sencillo. Pero no se conoce este valor, y no es posible conocerlo ya que son muchas las posibilidades de extraer muestras diferentes. Lo más adecuado es *estimarla con el valor muestral*, o *media muestral.* Se dice simplemente, en este caso, sea la media muestral igual a la media de las medias de la población y espérese estar en lo cierto. Entonces debe probarse la ecuación. Se lleva a cabo esto con el error estándar.

Se aplica una argumentación análoga a la desviación estándar de toda la población (de los puntajes originales). No se conoce y probablemente nunca pueda conocerse. Pero puede estimarse con la desviación estándar calculada a partir de la muestra. Nuevamente se dice, en efecto, sea la desviación estándar de la muestra igual a la desviación estándar de la población. Se sabe que probablemente no tienen el mismo valor, pero también se sabe, si el muestreo ha sido aleatorio, que probablemente no difieran mucho.

En el capítulo 11 se utilizó la desviación estándar de la muestra como un sustituto de la desviación estándar de la población, en la fórmula para el *error estándar de la media:*

$$SE_M = \frac{SD}{\sqrt{N}}$$

A esto también se le denomina *error de muestreo.* Así como la desviación estándar es una medida de la dispersión de los puntajes originales, el error estándar de la media es una medida de la dispersión de la distribución de las medias de las

muestras. *No es* la desviación estándar de la población de puntajes individuales si, por ejemplo, pudiera probarse cada miembro de la población y calcular la media y la desviación estándar de esta población.

UNA DEMOSTRACIÓN POR EL MÉTODO DE MONTECARLO

Para tener material con el que trabajar, ahora se recurre a la computadora y a lo que se denomina métodos de Montecarlo. Los métodos de Montecarlo son métodos de simulación manejados por computadora, diseñados para obtener soluciones a problemas matemáticos, estadísticos, numéricos e incluso verbales utilizando procedimientos aleatorios y muestras de números aleatorios.

Usualmente asociados con problemas matemáticos cuyas soluciones son intratables, los métodos de Montecarlo se han extendido a la "comprobación" de características estadísticas de muestras de poblaciones grandes. Por ejemplo, las consecuencias de violar las suposiciones detrás de las pruebas estadísticas de significancia pueden estudiarse efectivamente mediante la simulación de distribuciones estadísticas con números aleatorios y la introducción de violaciones a las suposiciones en el procedimiento para estudiar las consecuencias.

En las ciencias del comportamiento, los procedimientos de Montecarlo son usualmente estudios empíricos de modelos estadísticos y de otros tipos que utilizan los números aleatorios generados por la computadora para ayudar a simular los procesos aleatorios necesarios para estudiar los modelos. En cualquier caso, ahora se utiliza una forma elemental de Montecarlo para probar un teorema bastante importante de estadística y para investigar la variabilidad de medias y la utilización del error estándar de la media. También se desea establecer una base para la comprensión de la computadora al estudiar procesos aleatorios.

Procedimiento

Se escribió un programa de computadora para generar 4,000 números aleatorios uniformemente distribuidos entre 0 y 100 (de tal forma que cada número tiene la misma probabilidad de ser "extraído" en 40 conjuntos de 100 números cada uno, así como para calcular varios estadísticos con los números. Considérese este conjunto de 4,000 números como la población, o U. La media de U es 50,33 (por medio de un cálculo real de computadora), y la desviación estándar es 29,17. Se desea calcular esta media a partir de muestras extraídas al azar de U. Por supuesto, en una situación real usualmente no se conoce la media de la población. Una de las virtudes de los procedimientos de Montecarlo es que puede conocerse lo que usualmente se desconoce.

Cinco de los 40 conjuntos de 100 números fueron extraídos al azar. (Los conjuntos extraídos fueron 5, 7, 8, 16 y 36.Se calcularon las medias y las desviaciones estándar de los cinco conjuntos. Asimismo, se calcularon los cinco errores estándar de la media. En el cuadro 1 se muestran estos estadísticos. Se quiere proporcionar una noción intuitiva acerca de lo que es el error estándar de la media y posteriormente se quiere mostrar cómo es utilizado.

MEDIAS, DESVIACIONES ESTÁNDAR Y ERRORES ESTÁNDAR DE LA MEDIA, CINCO MUESTRAS DE 100 NÚMEROS ALEATORIOS (DE 0 A 100)[A]

	Muestras				
	1	*2*	*3*	4	*5*
M:	*53.21*	*49.64*	*51.37*	*49.02*	*55.51*
SD:	*29.62*	*27.91*	*29.83*	*26.72*	*29.23*
SE:	*2.96*	*2.79*	*2.98*	*2.67*	*2.92*

[A] Estadísticos de la población: M = 50,33; SD =29,17; N = 4,000

Figura 1

Primero, se calcula la *desviación estándar de esta muestra de medias*. Si simplemente se trata a las cinco medias como puntajes ordinarios y se calcula la media y la desviación estándar, se obtiene: M_1 = 51,75; *SD* = 2,38. La media de los 4,000 puntajes es 50,33. Cada una de las cinco medias es una estimación muestral de esta media de la población. Obsérvese que tres de ellas, 49,64, 51,37 y 49,02, se aproximan bastante a la media de la población, y que dos de ellas, 53,21 y 55,51 difieren bastante de la media de la población. Así pues, parece que tres de las muestras proporcionan buenos estimadores de la media de la población y que dos no lo hacen, ¿o si?

La desviación estándar de 2,38 es *semejante al* error estándar de la media. (No es, por supuesto, el error estándar de la media, ya que se ha calculado solamente a partir de cinco medias.) Supóngase que solamente se ha extraído una muestra (la primera, con M = 53,21 y SD = 29,62), lo que es la situación usual en la investigación, y que se calculó el error estándar de la media:

$$SE_M = \frac{SD}{\sqrt{N}} = \frac{29.62}{\sqrt{N}} = 2.96$$

Este valor es una estimación de la desviación estándar de las *medias* de la población de muchas muestras de 100 casos, cada una de las cuales se ha extraído al azar de la población. La población tiene 40 grupos y por ello 40 medias. (Naturalmente, éstas no son muchas medias.) La desviación estándar de estas medias es realmente 3,10. El SE_M calculado con la primera muestra, entonces, se aproxima al valor de esta población: 2,96 como una estimación de 3,10.

En el cuadro 1 se muestran los cinco errores estándar de la media (tercer renglón). Fluctúan muy poco, de 2,67 a 2,98, aun cuando las medias de los conjuntos de 100 puntajes varían considerablemente. La desviación estándar de 2,38 calculada a partir de las cinco medias es solamente una estimación justa de la desviación estándar de la población de medias. Aún así, es una estimación. La cuestión interesante e importante es que el error estándar de la media, que es una estimación "teórica", calculado a partir de los datos de cualquiera de los cinco grupos, es una estimación exacta de la variabilidad de las medias de muestras de la población.

Para reforzar estas ideas, ahora se considerará otra demostración de Montecarlo de mucha mayor magnitud. El programa de computadora utilizado para producir los 4,000 números aleatorios del ejemplo discutidos anteriormente fue usado para producir 15 conjuntos más de 4,000 números aleatorios cada uno, uniformemente distribuidos entre 0 y 100. Esto es, se generó un total de 80,000 números aleatorios en 20 conjuntos de 4 000 cada uno. La media teórica, de nueva cuenta, de los números entre 0 y 100 es 50. Considérese cada uno de los 20 conjuntos como una muestra de 4,000 números. En el cuadro 2 se muestran las medias de los 20 conjuntos.

MEDIAS DE 20 CONJUNTOS DE 4,000 NÚMEROS ALEATORIOS GENERADOS POR COMPUTADORA (DE 0 A 100)[a]

50,3322	49,9447	50,1615	50,0995
50,1170	*49,5960*	51,0585	51,1450
49,8200	49,3175	*49,5822*	50,6440
49,8227	*49,9022*	*49,7505*	49,8437
49,5875	50,6180	*50,0990*	49,3605

[a] Media de las Medias = 50,0401; desviación estándar de las medias = 4,956; error estándar de la media, primera muestra = ,4611.

Figura 2

Las 20 medias se acumulan muy estrechamente alrededor de 50: la más baja es 49,3175, la más alta es 51,1450 y la mayor parte de ellas se aproximan a 50. La media de las veinte medias es 50,0401, claramente muy próxima a la esperanza teórica de 50. La desviación estándar de las veinte medias es ,4956. La desviación estándar de la primera muestra de 4,000 casos (ver la nota a, al pie del cuadro 1) es

29,1653. Si se utiliza esta desviación estándar para calcular el error estándar de la media, se obtiene: SE_M = 29,1653 4000 =,4611. Obsérvese que esta estimación del error estándar de la media se aproxima a la desviación estándar calculada de las 20 medias. No se comete error al utilizar esto para evaluar la variabilidad de las medias de muestras de 4,000 números aleatorios. Es claro que las medias de muestras grandes son estadísticos altamente estables y los errores estándar son buenas estimaciones de su variabilidad.

Generalizaciones

Ahora pueden hacerse tres o cuatro generalizaciones de gran utilidad en la investigación. Una, las medias de las muestras son estables en el sentido de que son mucho menos variables que las mediciones a partir de las cuales son calculadas. Esto es, por supuesto, verdad por definición. Las varianzas, desviaciones estándar y los errores estándar de la media son inclusive más estables; fluctúan en márgenes relativamente estrechos. Aún cuando las medias de las muestras del ejemplo variaron tanto como cuatro o cinco puntos, los errores estándar fluctuaron en no más de un punto y medio. Esto significa que puede tenerse considerable fe en el hecho de que las estimaciones de las medias de muestras estarán muy próximas a la media de una población de tales medias. Y la ley de los números grandes afirma que mientras más grande sea el tamaño de la muestra, probablemente los estadísticos estarán más próximos a los valores de la población.

Una pregunta difícil para los investigadores es: ¿Siempre se cumplen estas generalizaciones, especialmente con muestras no aleatorias? La validez de las generalizaciones depende del muestreo aleatorio. Si el muestreo no es aleatorio, en realidad no es posible saber si se cumple la generalización. Sin embargo, muchas veces se tiene que actuar como si se pensara que se cumplen, incluso con muestras no aleatorias. Por fortuna, si se es cuidadoso respecto al estudio de los datos para detectar la idiosincrasia sustancial de la muestra, es posible utilizar con provecho la teoría. Por ejemplo, las muestras pueden comprobarse para esperanzas fácilmente verificadas. Si se espera más o menos el mismo número de machos que de hembras en una muestra, o proporciones conocidas de jóvenes y viejos o de republicanos y demócratas, es sencillo contar tales números. Hay expertos que insisten en el muestreo aleatorio como una condición para la validez de la teoría, y tienen razón. Pero si la teoría esta prohibida con muestras no aleatorias, entonces tendrían que abandonarse muchas de las utilizaciones de la estadística e inferencias. La realidad es que aparentemente la estadística funciona muy bien aun con muestras no aleatorias, suponiendo que el investigador conoce las limitaciones de tales muestras, incluso se tiene más cuidado que al trabajar con muestras aleatorias y los estudios se efectúan dos veces.

Teorema del límite central

Antes de estudiar la utilización verdadera del error estándar de la media, debe observarse, aun cuando sea brevemente, una generalización extremadamente importante concerniente a las medias: *Si las muestras son extraídas al azar de una población, las medias de las muestras tenderán a distribuirse normalmente*. Mientras más grandes sean las *n*, más verdadero será este hecho. Y la forma y clase de distribución de la población original no importa. Esto es, la distribución de la población no tiene que ser normal.

Por ejemplo, la distribución de los 4,000 números aleatorios referidos es rectangular, ya que los números están distribuidos uniformemente. Si el teorema del límite central es válido empíricamente, entonces las medias de cada uno de los 40 conjuntos de 100 puntajes deben estar más o menos distribuidas de manera normal.

De ser así, éste es un hecho notable. Y es así, aunque una muestra de 40 muestras difícilmente es suficiente para mostrar muy bien la tendencia. En consecuencia, se generaron en la computadora tres poblaciones más de 4,000 diferentes números aleatorios distribuidos de manera uniforme, divididos en 40 subconjuntos de 100 números cada uno. Se estimaron las medias calculadas para los 4 x 40 = 160 subconjuntos de 100 números cada uno y se colocaron en una distribución. En la figura 1 se muestra un polígono de frecuencias de las medias. Puede observarse que las 160 medias se asemejan bastante a la forma acampanada de la curva normal. Aparentemente "funciona" el teorema del límite central. Y no olvidarse de que esta distribución de medias se obtuvo de una distribución rectangular de números.

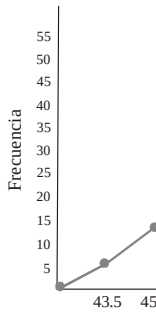


Figura 3

¿Por qué tomarse tanta molestia? ¿Por qué es importante mostrar que las distribuciones de medias se aproximan a la normal? Una gran parte del análisis de datos trata con medias, y si están distribuidas normalmente, entonces es posible utilizar las propiedades conocidas de la curva normal para interpretar los datos obtenidos en la investigación. El saber que alrededor del 96% de las medias caerán

entre dos desviaciones estándar (errores estándar) arriba y abajo de la media es una información valiosa, ya que un resultado obtenido puede valorarse contra las propiedades conocidas de la curva normal.

ERROR ESTÁNDAR DE LAS DIFERENCIAS ENTRE MEDIAS

Una de las estrategias más frecuentes y útiles en la investigación es comparar medias de muestras. A partir de las diferencias en las medias se infieren los efectos de las variables independientes. También cualquier combinación lineal está regida por el teorema del límite central. Esto es, dadas muestras lo suficientemente grandes, las diferencias en las medias estarán distribuidas en forma normal. (Una combinación lineal es cualquier ecuación de primer grado, p. ej., $Y = M_1 - M_2$. $Y = M_1^2 - M_1$ no es lineal.) En consecuencia, puede utilizarse la misma teoría que se usa con las medias en las diferencias entre medias.

Supóngase que se han asignado aleatoriamente 200 sujetos a dos grupos, cada uno de 100 sujetos. A un grupo se le muestra una película sobre relaciones intergrupales, por ejemplo, y nada al otro grupo. A continuación, se aplica a ambos grupos una prueba de actitudes. El puntaje medio del grupo A (que vio la película) es 110 y el puntaje medio del grupo B (que no vio la película) es 100. El problema es: ¿Es la diferencia de 10 unidades una diferencia "real", una diferencia estadísticamente significativa? ¿O es una diferencia que pudo haber surgido al azar, más de cinco veces en 100, por decir algo, o alguna otra cantidad, cuando realmente no existe ninguna diferencia?

Si análogamente se crean muestras dobles de 100 cada una y se calculan las diferencias entre las medias de esas muestras, y se lleva a cabo el mismo procedimiento experimental, ¿se obtendrá de manera coherente esta diferencia de 10? De nuevo, se utiliza el error estándar para evaluar las diferencias, pero esta vez se tiene una *distribución muestral de diferencias entre las medias.* Es como si se tomara cada $M_i - M_j$ se considerara como una X. Entonces las varias diferencias entre las medias de las muestras se consideran como las X de una nueva distribución. Para cualquier promedio, la desviación estándar de esta distribución muestral es *semejante* al error estándar. Pero este procedimiento es sólo para ilustración; en la realidad no se hace esto. Aquí, otra vez, se estima el error estándar a partir de los primeros dos grupos, A y B, utilizando la formula:

$$EE_{MA} - M_B = \sqrt{EE_{M^2} - EE_{M_{B^2}}}$$

Donde SE_{MA}^2 y SE_{MB}^2 son los errores estándar al cuadrado de los grupos A y B, respectivamente, como se había planteado previamente.

Supóngase que el experimento se llevó a cabo con cinco dobles grupos, es decir, con 10 grupos, dos cada vez. Las cinco diferencias entre las medias fueron 10, 11, 12, 8, 9. La media de estas diferencias es 10; la desviación estándar es 1,414. Este 1,414 nuevamente es *semejante* al error estándar de la distribución muestral de las diferencias entre las medias, en el mismo sentido que el error estándar de la media en la discusión precedente. Ahora, si se calcula el error estándar de la media para cada grupo (haciendo que las desviaciones estándar para los dos grupos sean $SD_A = 8$ y $SD_B = 9$) se obtiene:

$$EE_{M_B} = \frac{DE_B}{\sqrt{n_B}} = \frac{9}{\sqrt{100}} = 9 \qquad EE_{M_A} = \frac{DE_A}{\sqrt{n_A}} = \frac{8}{\sqrt{100}} = 8$$

Con la siguiente ecuación se calcula el error estándar de las diferencias entre las medias:

$$EE_M - M_B = \sqrt{EE_{M_B 2} + EE_{M_B 2}} = \sqrt{(.8)^2 + (.9)^2} = \sqrt{.64 + .81} = \sqrt{1.45} = 1.20$$

Ahora que se tiene el 1,20, ¿que se hace con él? Si los puntajes de los dos grupos han sido elegidos de un tabla de números aleatorios y no hubo condiciones experimentales, podría esperarse que no hubiera diferencia entre las medias. Pero se ha aprendido que siempre hay diferencias de pequeñez relativa debido a factores fortuitos. Estas diferencias son aleatorias. *El error estándar (SE) de las diferencias entre las medias es una estimación de la dispersión de estas diferencias.*

Pero es una medición de estas diferencias que es una estimación para toda la población de tales diferencias. Por ejemplo, el error estándar de las diferencias entre las medias es 1,20. Esto significa que, solamente por la casualidad, alrededor de la diferencia de 10 entre M_A y M_B habrá fluctuaciones aleatorias, ahora 10, luego 10,2, después 9,8, etcétera. Solamente rara vez las diferencias excederán, por ejemplo, 13 ó 7 (alrededor de tres veces el SE). Otra forma de plantear esto es decir que el error estándar de 1,20 indica los límites (si se multiplica el 1,20 por el factor apropiado) más allá de los cuales probablemente no pasarán las diferencias muestrales entre las medias.

¿Qué tiene que ver todo esto con el experimento? Es precisamente aquí que se evalúan los resultados experimentales. El error estándar de 1,20 fluctuaciones aleatorias estimadas. Ahora, $M_A - M_B = 10$. ¿Pudo haber surgido esto debido a la casualidad, como resultado de fluctuaciones aleatorias como las recién descritas? Por ahora debe estar casi claro que esto no puede ser, excepto bajo ciertas circunstancias poco comunes. Se evalúa esta diferencia de 10 comparándolo con la

estimación de las fluctuaciones aleatorias o fortuitas. ¿Es una de ellas? La comparación se lleva a cabo mediante la razón *t, o* prueba *t.*

$$t = \frac{M_A - M_B}{EE_{MA} - M_B} = \frac{110 - 100}{1,20} = \frac{10}{1,20} = 8,33$$

Esto significa que la diferencia medida entre M_A y M_B debería estar a 8,33 desviaciones estándar de una media hipotética de cero (diferencia cero, ninguna diferencia entre las dos medias).

No se tendría ninguna diferencia, desde el punto de vista teórico, si los sujetos fueran totalmente azarosos y si no hubiera manipulación experimental. Se tendrían, en efecto, dos distribuciones de números aleatorios a partir de las cuales podrían esperarse sólo fluctuaciones fortuitas. Pero aquí se tiene, comparativamente, una gran diferencia de 10, en relación con una insignificante 1,20 (la estimación de las desviaciones aleatorias). Es definitivo que algo más allá del azar está sucediendo aquí. Y este algo es en realidad lo que se está buscando. Es, presumiblemente, el efecto de la película, o el efecto de la condición experimental, otras condiciones que, por supuesto, han sido controladas lo suficiente.

Representa *una población de diferencias entre medias* con una media de cero y una desviación estándar de 1.20. (La media se fija en cero porque se supone que la media de todas las diferencias de medias es cero.) ¿Dónde debe ubicarse la diferencia de 10 sobre la línea horizontal del diagrama? Para responder esta pregunta, primero debe convertirse el 10 a unidades de desviación estándar (o de error estándar).

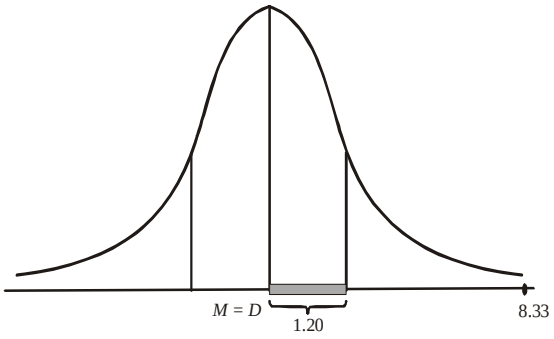


Figura 4

Esto se lleva a cabo dividiendo entre la desviación estándar (error estándar), que es 1,20: 10/1,2 = 8,33. Pero esto es lo que se obtuvo cuando se calculó la razón t. Es, entonces, sencillamente la diferencia entre M_A y M_B, 10, expresada en unidades de desviación estándar (error estándar). Ahora puede ubicarse en la línea horizontal del diagrama. Obsérvese el punto que se encuentra alejado a la derecha. Resulta

claro que la diferencia de 10 es una desviación. Se encuentra tan alejada, de hecho, que a lo mejor no pertenece la población en cuestión. Brevemente, la diferencia entre M_A y M_B es de significancia estadística; tan significativa que explica lo que Bernoulli denominó "certidumbre moral". Es difícil que una diferencia tan grande, o desviación de la esperanza casual, pueda atribuirse a la casualidad. En realidad las probabilidades son mayores que un billón a uno. *Puede* suceder. Pero es mínima la probabilidad de que ocurra.

Así es el error estándar y su utilización. Los errores estándar de otros estadísticos se usan de la misma manera. Una herramienta muy importante y útil. Es un instrumento básico en la investigación contemporánea. Evidentemente, sería difícil imaginar la metodología moderna de la investigación, e imposible imaginar la estadística moderna, sin el error estándar. Como una de las claves de la inferencia estadística, no puede exagerarse su importancia. Gran parte de la inferencia estadística cae bajo la familia de fracciones compendiadas por la fracción:

$$\frac{\text{Estadístico}}{\text{Error estándar del estadístico}}$$

INFERENCIA ESTADÍSTICA

Inferir es derivar una conclusión a partir de premisas o de la evidencia. *Inferir estadísticamente* es derivar conclusiones probabilísticas a partir de premisas probabilísticas. Se concluye probabilísticamente, es decir, a un nivel especificado de significancia. Se infiere, en términos probabilísticos, si un experimento resulta desviado de la esperanza fortuita, si la hipótesis nula no es "verdadera", que está funcionando una influencia "real". Si, en los métodos del experimento, $M_A > M_B$ y $M_A \neq M_B$ o H_1 es "verdadera" y H_o no es "verdadera", se infiere que el método A es "superior" al B, siendo "superior" aceptado en el sentido definido en el experimento.

Otra forma de inferencia, comentada a lo largo del capítulo sobre muestreo, es aquélla obtenida a partir de una muestra de la población. Ya que, por ejemplo, el 55% de una muestra aleatoria de 2,000 personas en Estados Unidos afirma que votará por un cierto candidato presidencial, se infiere que, si se le preguntara a toda la población del mencionado país, responderá de forma similar. Ésta es una inferencia muy grande. Uno de los peligros graves en la investigación, o tal vez debiera decirse de cualquier razonamiento humano, es el salto inferencial a partir de datos muestrales hacia hechos de la población. Constantemente se efectúan saltos inferenciales que no son de tamaño medio: en política, economía, educación y otras áreas de gran importancia. Por ejemplo, si el gobierno reduce los gastos, la inflación decrecerá. Si se utilizan máquinas para enseñar, los niños aprenderán más. Pero los científicos,

también, efectúan saltos inferenciales, a menudo algunos muy grandes, con una diferencia importante. El científico es (o debería ser) consciente de que está realizando tales saltos y que siempre son arriesgados.

En resumen, puede afirmarse, que la estadística permite al científico probar indirectamente hipótesis sustantivas, capacitándolo para probar de manera directa hipótesis estadísticas (en caso de ser posible probar de esa forma cualquier cosa). En este proceso, se usan las hipótesis nulas, hipótesis escritas por casualidad. Se prueba la "verdad" de hipótesis sustantivas al someter las hipótesis nulas a pruebas estadísticas con base en el razonamiento probabilístico. Entonces pueden hacerse inferencias apropiadas. Es claro que el objetivo de todas las pruebas estadísticas es probar lo justificable de las inferencias.

CAPÍTULO IV

EL INFORME CIENTÍFICO

EL INFORME CIENTÍFICO

1. COMUNICACIÓN DE LA INVESTIGACIÓN

La investigación culmina necesariamente con un informe que se debe redactar en cumplimiento de la obligación que ha asumido el investigador con la comunidad científica de dar a conocer los resultados de sus investigaciones y de los hallazgos realizados. Pero ¿dónde se origina esta obligación? La obligación de comunicar a la comunidad científica se origina en el momento en el que el investigador parte de conocimientos previos para identificar problemas de investigación. Y estos conocimientos previos son patrimonio de la humanidad, que si bien es cierto están a disposición de todos los hombres, han servido de punto de partida para producir nuevos conocimientos. Estos nuevos conocimientos no pertenecen por entero a quien los produjo, sino son el resultado de una cadena de aportes que en el tiempo han venido realizando investigaciones precedentes. Aquí es oportuno recordar lo que decía Newton "Yo sólo veo parado sobre los hombros de gigantes", con lo que quería demostrar que los nuevos conocimientos en Newton se concatenaban con los conocimientos previos producidos por investigadores que lo antecedieron. Es la confesión de parte de un investigador honesto que atribuye sus éxitos en la investigación a los conocimientos previos acumulados por sus antecesores ante quienes, con humildad, se empequeñece. Lo que quería destacar Newton es la concatenación que existe entre los nuevos conocimientos con los conocimientos previos, con los que describía, con gran acierto, la naturaleza de la ciencia. En efecto, nada nuevo aportan las investigaciones que no hayan sido materia de preocupación de científicos que precedieron al investigador en el estudio el tema. Por eso se dice que todo científico tiene la obligación de comunicar sus hallazgos para permitir que avance la ciencia, pues los hallazgos a los que ha llegado tienen su punto de partida en los conocimientos previos que pertenecen a la comunidad científica. Pero al mismo tiempo, los hallazgos a los que ha llegado servirán, en el futuro, de antecedentes para las nuevas investigaciones que acometan nuevos hombres de ciencia. Nada nuevo aportan las investigaciones que no hayan sido previamente materia de preocupación de otros científicos.

Para redactar el informe científico, existen algunos criterios básicos que cumplen una función orientadora. En primer lugar se debe preparar un esquema, estructura o modelo según el cual se debe redactar el informe. Sobre el particular existe una

diversidad de opiniones de los autores sin embargo, subyace en todas ellas, una lógica aceptable que orienta los esfuerzos por redactar un buen informe. En definitiva, el investigador debe asumir un determinado esquema según el cual debe redactar su informe.

Es muy conveniente que el informe comience por un resumen de su contenido, lo que facilita la comprensión de la temática que aborda. Los investigadores que revisan la literatura, al leer el resumen o *abstract* se forman una idea acerca del contenido del informe y si lo consideran pertinente, leerán todo el informe o lo dejarán para una oportunidad posterior si se enteran, por la lectura del resumen, que el contenido del informe no satisface sus necesidades. Por eso se recomienda que el resumen se redacte, además de la lengua en que está escrito el informe, en una lengua extranjera, preferentemente el inglés, por ser ésta, la que más difusión tiene en el campo de la ciencia

Después del resumen se redacta la introducción, que viene a constituir una referencia muy general con respecto al contenido del informe. Muchas veces, cuando la introducción está bien elaborada, es suficiente para que el lector se forme una idea del contenido de todo el informe. En esta introducción se debe hacer referencia al contenido de los distintos capítulos del informe.

El cuerpo del informe está constituido por los capítulos que contiene. Un primer capítulo debe estar referido al planteamiento del estudio que se llevó a cabo. Como su nombre lo dice, en este capítulo se debe contextualizar la investigación y permitir que el lector se ubique y comprenda, en una primera aproximación, acerca de lo que trata el trabajo. Por eso es preciso que en este primer capítulo se plantee y se explique el problema, los objetivos, la hipótesis de la investigación y, lo que es más importante, que se identifiquen, con claridad meridiana, las variables que se han estudiado y dar cuenta de la forma cómo han variado, si se las ha considerado como dicotomías o como politomías. También es necesario dar cuenta de los criterios empleados para clasificar las variables. La clasificación de las variables es muy importante explicarlas porque permite un conocimiento, a mayor profundidad, de su naturaleza y de sus características. Así el lector puede saber si las variables varían en dos, tres o más valores, si ha sido justificado el uso de métodos cualitativos o cuantitativos, o si las variables han sido manipuladas o simplemente observadas. La presentación de la clasificación de las variables también es importante porque ayuda a entender las razones por las que se ha adoptado determinada estrategia para la prueba de hipótesis. Es conveniente incluir en este capítulo la justificación del estudio y sus respectivos alcances y limitaciones. Este primer capítulo tiene la virtud de encuadrar el estudio y servir de guía para la comprensión del contenido del resto del informe.

Un segundo capítulo debe estar dedicado a exponer las bases teóricas sobre las que se sustenta la investigación. Aquí es conveniente iniciar el capítulo haciendo un balance del estado en el que se encuentran las investigaciones relacionadas con el tema de estudio. Algunos autores lo denominan "estado de la cuestión" y en efecto, este capítulo se justifica en tanto y en cuanto permite que el autor de la investigación dé cuenta del trabajo de revisión bibliográfica que ha realizado. La revisión de la literatura –para ello es imprescindible el dominio de una segunda lengua– es importante porque permite explicar que la investigación realizada no se encuentra aislada, sino que se entronca con investigaciones previas, referidas al tema. Se debe dar cuenta de las investigaciones precedentes, precisamente para demostrar que el proceso de producción de nuevos conocimientos es una secuencia concatenada de los trabajos previos con los nuevos. Los antecedentes deben referirse no sólo a los estudios previos directamente relacionados con el tema de estudio, sino también a estudios afines que tengan alguna vinculación con el tema. A este respecto cabe mencionar que muchos investigadores noveles afirman que los estudios que han realizado no tienen precedentes en la literatura especializada. Nada más falso. A estas alturas del avance científico tecnológico alcanzado por la humanidad, no hay, en estricto sentido, "nada nuevo bajo el Sol". Es aquí que el investigador se da cuenta que, aún de manera tangencial, otros investigadores ya han tocado, de alguna manera, el tema acerca del cual informa. La tarea de ubicar antecedentes de la investigación es ardua e intensa y debe iniciarse a nivel local, luego generalizarse a nivel nacional, para luego dar cuenta de los antecedentes que tendría el estudio a nivel internacional.

En este acápite el investigador afirma que sigue una tradición que viene de mucho antes y demuestra que tiene respaldo académico que le permite reclamar, para su investigación, el valor que pudiera corresponderle. En otras palabras, con este acápite el investigador demuestra que no ha estado sólo al investigar la temática que expone, sino que existen otros investigadores que también se ocupan del tema. Con este acápite, además, es posible establecer los límites a los que la investigación sobre el tema ha llegado. Por eso se dice "estado de la cuestión", porque proporciona el balance acerca del progreso del conocimiento, señalando los logros alcanzados e identificando los aspectos que la ciencia aún no explica. Por ejemplo, si el investigador ha estudiado la problemática de la calidad de la formación profesional universitaria, debe explicar que muchas investigaciones precedentes se han ocupado de esta temática y reportar que, hasta la fecha, la investigación sobre el tema ha permitido identificar una cierta cantidad de factores causales que estarían influyendo en el fenómeno que estudia. Este acápite se justifica porque permite al autor de la investigación dar cuenta de la revisión bibliográfica que ha realizado.

Como consecuencia de la revisión de la literatura, el investigador ha recolectado mucha información valiosa y debe presentarla en forma coherente y ordenada. Por ejemplo si ha revisado informes científicos a nivel local, nacional o internacional, debe presentar la información siguiendo este orden, tratando de organizar sistemáticamente la información que posee.

En este capítulo también se debe considerar un segundo acápite referido a las bases teóricas sobre las cuales se ha realizado la investigación. Con esto el investigador reporta acerca del soporte teórico que ha tenido en cuenta para realizar la investigación de cuyos resultados informa. Toda investigación debe tener su respectivo soporte teórico que permita entroncar los resultados de la investigación en el cuerpo de conocimientos previos que han acumulado los científicos, a lo largo de la historia.

La revisión de la literatura permite elaborar una reseña de las principales ideas que sirvieron de base teórica a la investigación realizada. Esta reseña debe estar ordenada en una estructura jerarquizadas de conceptos por cada variable estudiada. Se debe evitar lo que muchas veces se ha observado: la repetición o la copia literal de la información teórica hallada al respecto, sin orden ni coherencia. Aquí, el investigador que redacta el informe tiene la oportunidad de demostrar su erudición en el tema y no sólo eso, debe demostrar su formación académica y el rigor con el que trata la información científica. En este acápite el investigador debe hacer las citas textuales necesarias, los parafraseos de ideas relevantes o los comentarios sagaces y oportunos, todo ello sustentado en las referencias bibliográficas que sirven de fuente a las ideas que se está exponiendo. También en este acápite el investigador debe asumir una posición con respecto a las distintas perspectivas teóricas que pudiera haber hallado en sus lecturas y debe manifestar sus discrepancias o coincidencias con las bases teóricas que ha revisado y tratar de elaborar un cuerpo teórico capaz de sustentar la investigación realizada. Él debe tratar de construir una teoría que sustente sus investigaciones ensayando una propuesta coherente que sustente el estudio.

Conviene hacer hincapié en este punto, a propósito de que muchas veces, los investigadores noveles lo que hacen es recolectar y repetir la información teórica al respecto, sin organizar la información en forma coherente y ordenada. Y lo que es más grave, repiten o copian la teoría que muchas veces ni se relaciona con el tema.

Finalmente, en este capítulo, y como consecuencia de lo anteriormente asumido, se deben consignar las definiciones conceptuales de los constructo que le han servido al investigador para asumir una posición teórica al respecto, o bien, redefinir o asumir significaciones particulares de los conceptos que se manejan en el estudio, cuando se usen constructos o terminología que en el ambiente académico se emplea

en otro contexto o en otros campos semánticos. En otras palabras, se trata de elaborar un glosario de términos *ah doc* a la investigación que se ha realizado. Por ejemplo, si el investigador, para referirse al trabajo que desempeña un profesional, emplea el constructo *práctica profesional*, debe explicar que esta frase se emplea para referirse al ejercicio profesional que desarrolla todo profesional, es decir, que el investigador emplea esta frase en un contexto distinto al que se emplea en los procesos de formación profesional, en los que *la práctica profesional* se refiere a la ejercitación que debe realizar un estudiante, antes de graduarse, en las tareas propias de su profesión.

Un tercer capítulo debe estar referido al aspecto metodológico que ha servido para realizar la investigación. En este capítulo se debe informar acerca de la operacionalización de las variables, trabajo importante que debe realizar todo investigador. Luego debe dar cuenta del tipo de investigación que ha realizado, especificando los criterios que ha empleado para tipificar su investigación. También en este capítulo debe dar cuenta acerca de la población de estudio y para ello el investigador debe explicar todo lo relacionado con la población y con las técnicas de muestreo que ha utilizado.

En este capítulo debe explicar también la estrategia que ha aplicado para probar la hipótesis, es decir, debe explicar las previsiones adoptadas para poner a prueba las hipótesis con los hechos de la realidad. Aquí debe explicar si su estrategia ha sido experimental o cuasi experimental, si ha trabajado con una o más variables independientes, y si esta estrategia le ha permitido controlar exitosamente las variables intervinientes.

Otro acápite importante de este capítulo es el referido a los instrumentos que ha utilizado el investigador para recolectar sus datos. Estos instrumentos pueden haber sido tomados de la literatura, es decir instrumentos elaborados por otros autores y que se han aplicado en esta investigación o bien dar cuenta del proceso en el que el propio investigador, al no encontrar o no hallarse satisfecho con los resultados que podría obtener recolectando datos con los instrumentos existentes, ha tenido que elaborarlos de modo *ex profeso* para la investigación cuyo informe está realizando. El investigador debe explicar que ha aplicado ciertos instrumentos para recolectar datos acerca de cada variable. No es posible que el investigador refiera instrumentos sin hacer alusión a la variable para la cual han servido.

En este capítulo, también debe dar cuenta acerca del proceso de la prueba de hipótesis, lo que por lo general se hace empleando las pruebas estadísticas disponibles, o bien empleando alguna metodología específica para el caso de las investigaciones cualitativas que en los últimos años se han puesto de moda.

En un capítulo final el investigador debe dar cuenta del trabajo de campo realizado. Este capítulo tal vez sea el más importante de todo el informe, porque es aquí donde el investigador presenta sus hallazgos y da cuenta del modo cómo ha arribado a ellos. En este punto el investigador explica las técnicas que ha utilizado, y en caso que la investigación se haya realizado empleando el método hipotético deductivo, debe dar cuenta de cómo ha adoptado las decisiones con respecto a la hipótesis nula, es decir, cómo ha tomado las decisiones con el empleo de la metodología de la prueba inversa.

Finalmente, el investigador debe redactar las conclusiones a las que los hechos le han conducido. Las conclusiones deben ser muy precisas y deben estar referidas a las decisiones que el investigador ha adoptado con respecto a las hipótesis. En las conclusiones debe establecer si la hipótesis ha sido aceptada o rechazada y también debe precisar acerca de las consecuencias que podrían producir las decisiones adoptadas.

Como consecuencia de las conclusiones a las que el investigador ha arribado, se generan las respectivas recomendaciones, para que los investigadores, que acometan nuevas investigaciones acerca del tema o pretendan replicarlo, las tengan en cuenta.

En páginas finales es conveniente colocar los anexos, que vienen a ser los documentos complementarios al informe y que proporcionan mayores detalles para su mejor comprensión. En los anexos pueden aparecer los instrumentos de acopio de datos, los datos más puntuales y pertinentes acerca de la estrategia diseñada para probar las hipótesis, o bien la información complementaria referida al proceso de reducción de la población a la muestra estudiada, entre otros aspectos.

A continuación se presenta la estructura de un posible informe de investigación.

ESTRUCTURA DEL INFORME FINAL

TÍTULO
SUMARIO
RESUMEN
INTRODUCCIÓN
CAPÍTULO I: PLANTEAMIENTO DEL ESTUDIO
1. Fundamentación y formulación del problema.
2. Objetivos.
3. Justificación.

4. Alcances.
5. Fundamentación y formulación de las hipótesis.
6. Identificación y clasificación de las variables.

CAPÍTULO II: MARCO TEÓRICO

1. Antecedentes de la investigación.
2. Bases teóricas.
3. Definición conceptual de términos.

CAPÍTULO III: METODOLOGÍA DE LA INVESTIGACIÓN

1. Operacionalización de las variables.
2. Tipificación de la investigación.
3. Estrategia para la prueba de hipótesis.
4. Población y muestra.
5. Instrumentos de recolección de datos.

CAPÍTULO IV: TRABAJO DE CAMPO Y PROCESO DE CONTRASTE DE LA HIPÓTESIS

1. Presentación, análisis e interpretación de los datos.
2. Proceso de la prueba de hipótesis.
3. Discusión de los resultados.
4. Adopción de las decisiones.

CONCLUSIONES

RECOMENDACIONES

BIBLIOGRAFÍA

- Bibliografía referida al tema.
- Bibliografía referida a la metodología de investigación.

ANEXOS

- Cuadro de consistencia.
- Instrumentos de recolección de datos.
- Cuadros y gráficos.
- Tablas de interpretación de datos.

2. EXTENSIÓN DEL INFORME

Otro aspecto que se debe tener en cuenta al redactar el informe de la investigación es lo relacionado con su extensión, que también tiene que ver con la profundidad o amplitud con la que se redacta dicho informe. En otras palabras, aquí se debe tener en cuenta la extensión del informe. Éste debe ser tal, que no sea ni tan escueto, que omita información valiosa, ni tan detallada que contenga información intrascendente. Un criterio que serviría para redactar un buen informe de investigación, en cuanto a su extensión, es que debe consignar la información suficiente y necesaria como para que otro investigador, con el informe que tiene entre manos, pueda replicar la investigación. Es decir, quien redacta un informe debe redactarlo pensando en proporcionar la información que se requiere y que sea relevante como para replicar dicho estudio. La extensión del informe dependerá, pues, de este criterio.

Un informe muy minucioso, que proporciona muchos detalles irrelevantes, distrae la atención del lector con información intrascendente. Por el contrario, un informe muy escueto, no proporciona datos acerca de los grandes aspectos de la investigación, tales como el problema, las hipótesis, las variables, o la estrategia empleada para probar las hipótesis.

3. ESTILO DE REDACCIÓN DEL INFORME

El informe científico debe redactarse empleando el lenguaje científico técnico, que no es otra cosa que el uso del lenguaje en función denotativa. Usar el lenguaje denotativamente es dar cuenta de la realidad haciendo referencia sólo a hechos observables, objetivables, demostrables. Al usar denotativamente el lenguaje, el autor no emite sus impresiones personales acerca del fenómeno al que se refiere, ni mucho menos sus opiniones o juicios personales. El investigador debe ser objetivo y sólo da cuenta de los hechos, sin emitir sus apreciaciones subjetivas acerca de los hechos.

En el aspecto gramatical, el autor debe redactar en modo impersonal, es decir sin hacer referencia a su persona: “se halló”, “se demostró”, “se encontró”, son fórmulas recomendables para redactar informes científicos. Otra alternativa es usar el plural de cortesía, es decir, en vez de decir “yo hallé”, “yo encontré”, “yo demostré”, se recomienda usar las formas: “hallamos”, “demostramos”, “encontramos”. Con el plural de cortesía, aun sabiendo que el autor del informe es una persona individual, ésta se mimetiza demostrando con ello humildad académica al usar el plural aun cuando él solo haya hecho los descubrimientos que reporta.

4. REFERENCIAS BIBLIOGRÁFICAS

Se hallan estandarizados algunos sistemas para realizar las referencias bibliográficas. No existe, ni debe existir tampoco, predilección por ninguna de ellas, pero lo que sí importa es que el autor, al redactar el informe científico, debe adoptar un estilo determinado y emplearlo coherentemente a lo largo de su informe. Es muy recomendable empelar, en todo el informe, un solo estilo con lo que se proporciona uniformidad y calidad académica al trabajo.

Por ejemplo, si el autor desea hacer referencia a un texto de Piaget, entonces, después de la cita colocará entre paréntesis lo siguiente:

(Piaget, 69: 1978)

69 hace referencia a la página de la que se extrajo la cita y 1978 es el año de edición del libro citado. Conviene poner, en la bibliografía, el año de edición para identificar el título del libro, pues es poco común que un autor publique más de un libro por año. O si se trata de una traducción o una reimpresión, con el dato del año se puede deducir el título del libro. Naturalmente, en la bibliografía, debe aparecer el libro referido con su asiento bibliográfico completo, de este modo, el lector podrá identificar el título del libro de Piaget al que se hace referencia con esta fórmula.

Esta es una forma práctica de hacer la referencia bibliográfica. Aquí no se necesita hacer citas al pie de página y la lectura puede hacerse de corrido y tener la referencia de la bibliografía puesta de modo escueto dentro del texto que se está leyendo.

La forma convencional de hacer referencias bibliográficas es mediante las citas al pie de página o al final del texto. En este caso, al final del texto citado, el autor debe colocar, en súper índice, el número correlativo de las citas. En forma correspondiente a los súper índices, al pie de página, y en letras más pequeñas que las del texto, debe aparecer la cita en los siguientes términos:

[1] Salkind, Neil J. (1999): Métodos de Investigación. México, Editorial Prentice Hall Hispanoamericana S.A., pág. 87

Como se aprecia en el ejemplo que antecede, se ha colocado una línea corta, de no más de dos centímetros de longitud, para indicar que lo que sigue corresponde a las citas al pie de página. A renglón seguido aparece el numeral 1, en súper índice, que corresponde al mismo número que aparece en el párrafo que se cita. Es obvio que las citas siguientes serán numeradas en forma correlativa. Es posible, que en una sola página aparezcan una o más citas; esto depende de la necesidad de citar que tiene el redactor del informe.

En seguida aparece el apellido del autor. Los nombres del autor van después de una coma, algunos autores, con la finalidad de destacar la información que proporcionan escriben, en mayúsculas, el apellido del autor citado. Esta es una práctica muy usada, pero para mantener la uniformidad del texto del informe que escribe, en este caso, el apellido del autor citado debe aparecer en altas y bajas, vale decir, sólo la letra inicial del apellido debe ir en mayúsculas.

Algunos libros son escritos por varios autores. En estos casos, si los autores son dos, se consignan los nombres y apellidos de los dos autores, como se puede apreciar en el siguiente ejemplo:

[1] JOBERG, Gideon y NETT, Roger, (1980): Metodología de la investigación Científica. México, Editorial Trillas, pág. 56.

En este ejemplo se puede apreciar una variante en la forma de presentar los apellidos de los autores: van en mayúsculas y así fácilmente se puede saber que los autores son Sjobereg y Nett.

En algunos casos, si los autores son tres, sólo se escriben los apellidos de los tres autores, del modo como se indica en seguida:

[1] FERRANDEZ, SARRAMONA Y TARIN, (1988): Tecnología Didáctica. Barcelona, Ediciones CEAC., pág. 93.

En otros casos, si los autores son tres o más de tres se escribe sólo el apellido y el nombre del autor principal y se coloca luego la frase “y colaboradores” para indicar que la obra fue escrita por varios autores, colaboradores del autor principal. Para hacer referencia a los colaboradores también se usa la forma latina “et al”, abreviatura de et alter, como se ilustra en el siguiente ejemplo:

[1] ARY, Donald, et al. (1982): Introducción a la investigación pedagógica. México, Editorial Interamericana, pág. 109.

Luego de una coma, y entre paréntesis, se escribe el año de publicación el libro, tal como se puede ver en los ejemplos que se comentan. Antecedido por dos puntos (:) va el título del libro, con altas y bajas, si el título es breve; pero si el título es extenso, sólo se escribe el mayúsculas la letra inicial de la primera palabra del título.

En seguida, y formando un solo párrafo, se consigna la ciudad en la que se imprimió el libro y, seguido de una coma, el nombre de la editorial. Luego de otra coma, se da cuenta de la página de la que se extrajo la cita, empleando la abreviatura “pág”. Si la cita ha sido extraída de varias páginas, se coloca la abreviatura “pp.”, para referir el intervalo de páginas de la que se extrajo la cita. Por ejemplo:

[1] Salkind, Neil J., (1999): Métodos de Investigación. México, Editorial Prentice Hall Hispanoamericana S.A., pp. 87-93

Todos estos datos conforman un solo párrafo de dos renglones, más o menos, por lo que es muy recomendable usar una sangría francesa de 4 ó 5 puntos, para destacar los datos que van en el primer renglón, como son el súper índice correlativo y el apellido del autor. Si no se usara la sangría francesa, y todo el texto apareciera alineado a la izquierda, se haría más difícil identificar el orden correlativo de las citas y el apellido del autor se perdería en el contenido del párrafo. Esta forma de diagramar las citas al pie de página se puede apreciar en los ejemplos que anteceden.

Los científicos utilizan con mucha frecuencia información proveniente de publicaciones periódicas, es decir artículos de revistas especializadas. Para citar artículos publicados en revistas u otras fuentes hemerográficas, la alternativa es la siguiente:

[1] Peñalosa Ramella, Walter, (2000): El problema de las competencias. En: Revista Peruana de Educación. Año I, Nº 5 Lima, Optimice Editores, pp. 5-28.

Para destacar la fuente, se subraya el nombre de la revista en la que apareció el artículo citado y luego se proporciona el asiento bibliográfico de la revista, indicando el año y el número de la revista, así como la ciudad y la editorial.

Con la aparición de los medios electrónicos que permiten acceso rápido y variado a la información, surge la necesidad de citar la información a la que se accede por estos medios. Estos textos tienen sus propios títulos y sus respectivos autores, de modo que en las referencias bibliográficas que se hagan se deben consignar estos datos. Es obvio que en la información obtenida por estos medios no aparece la editorial o la ciudad, aunque sí se puede saber el año de la publicación.

Sin embargo se observa en recientes informes científicos que en vez de hacer las referencias bibliográficas sólo se consigna la Página web o la dirección electrónica de la que se obtuvo la información. Ante estas circunstancias, conviene recomendar la siguiente manera de citar la información obtenida por medios electrónicos, señalando que existen diversas alternativas propuestas.

[1] Salas, Rubén Darío: Las élites rioplatenses y su representación en la categoría de gobierno despótico. Revista electrónica, Número 1, junio de 1999.

http://constitucion.rediris.es/revista/hc/uno/diario2.html.Párrafo 3.

Debido a que las páginas de Internet carecen de numeración, para facilitar la búsqueda, se debe consignar el número del párrafo, como se ha hecho en el presente caso.

LECTURAS COMPLEMENTARIAS

EL INFORME DE INVESTIGACIÓN

Orfelio León *

Si ha tenido la oportunidad de viajar fuera de su país es probable que haya notado con agrado que casi la totalidad de las señales de tráfico son iguales. Visto con los ojos de alguien que llega a la mayoría de edad en las postrimerías del siglo XX, ésta parece la única forma razonable de haberlo hecho. Sin embargo, fíjese que existen otros contraejemplos: simultáneamente, en unas islas europeas, de no poca importancia, en vez de conducir por la derecha lo hacen por la izquierda. Todo lo que parece razonable y eficaz no es siempre asumido culturalmente. Lo que unos ven como facilitación de las comunicaciones los otros pueden verlo como colonialismo cultural. La normalización sacrifica la identidad y la peculiaridad individual en aras de un mejor entendimiento del grupo. En el caso que nos ocupa, la comunicación científica, se aboga por la normalización.

Las modernas publicaciones periódicas han ido imponiendo unos formatos para los informes de investigación. Estas normas de publicación suelen figurar en las contraportadas de las revistas, así que si quiere satisfacer su curiosidad, no tiene más que pedir una revista científica la próxima vez que vaya a la biblioteca de su Facultad. Los formatos a los que estamos aludiendo hacen referencia a *qué* hay que contar, en *qué orden* hay que contarlo *y qué aspecto* debe tener una vez mecanografiado. En Psicología, tanto en España como en el resto de los países occidentales, las normas se basan en las acordadas por la Asociación Americana de Psicología (APA. *Publication Manual, 3.*ª edición, 1983). *Si* alguna vez tiene ocasión de consultar el referido manual, verá que no hay nada dejado al azar. Es incluso posible que se sienta violento ante la rigidez de unas medidas que al ser pensadas originalmente en pulgadas producen unas traducciones con un nivel de detalle de diezmilímetros. En cualquier caso, no se preocupe que no tenemos la intención de trasladarle el nivel de puntillosidad de la APA a sus trabajos de alumno de primer ciclo. Es suficiente con que ordene los informes de sus prácticas con arreglo a la *estructura de los informes de investigación* que le vamos a presentar:

* Tomado de: LEON, Orfelio: *Diseño de Investigaciones. Introducción a la lógica.* Madrid, McGraw-Hill, 1993. pp. 292-300.

TÍTULO

RESUMEN

INTRODUCCIÓN

MÉTODO

RESULTADOS

DISCUSIÓN

REFERENCIAS

APÉNDICES

Efectivamente, esta estructura ya la conoce. Es casi la misma que hemos utilizado desde el Capítulo 4 para presentar los ejemplos de los experimentos. Y la razón por la que lo hicimos así era para que se fuera familiarizando con ella. Ahora lo que necesitaba saber es que no era una elección caprichosa, sino que se trata de la misma estructura que se encontrará en las revistas científicas. Los comentarios que vamos a realizar a continuación van dirigidos a ayudarle a redactar cualquier tipo de informe de práctica que tenga que hacer, manteniendo presente el formato común al que nos acabamos de referir.

Título

Dedique una página entera a poner un título, su nombre y filiación. El título debe describir el trabajo de la manera más específica. La filiación del autor hace referencia a la institución en la cual trabaja: universidad, instituto o empresa. En su caso puede hacer referencia a la facultad, curso y grupo.

Con el objeto de poder identificar todas las páginas del mismo informe, es conveniente que en el pie de la página del título aparezca una frase que le resuma. Es recomendable que utilice sólo dos o tres palabras. Este resumen se utilizará como *encabezamiento* en todas las páginas del informe y deberá colocarlo junto al número de la página. Por ejemplo, este texto tiene como título "Diseño de investigaciones: Introducción a la lógica de la investigación en Psicología y Educación", y como encabezamiento: "Diseño de investigaciones".

Resumen

Los resúmenes aparecen al *principio* del informe. Deben tener alrededor de 150 palabras. Se debe señalar en ellos lo más fundamental de los contenidos de la introducción, método, resultados y discusión. Si se utiliza algún procedimiento, aparato o test novedoso, conviene resaltarlo. El objetivo al hacer un resumen es que sea lo suficientemente claro y atractivo como para que alguien que duda si leer el trabajo,

al conocer el resumen, se decida a hacerlo. Se aconseja redactarlo cuando ya se haya completado el resto de los apartados.

En las revistas de habla no inglesa, junto al resumen aparece un resumen en inglés *(abstract)*.

Introducción

En primer lugar mentalícese de que la redacción de un informe de investigación no es el lugar más adecuado para descubrir nuevos valores de la literatura. Procure ser escueto, preciso, simple y por encima de todo: *claro.* Juan Benet o el último Camilo José Cela serían sus peores ejemplos. Azorín o Baroja serían unos excelentes modelos.

Comience haciendo una *descripción general del problema* que se aborda en su trabajo. Procure utilizar los mismos términos que usan los autores que ha consultado. No sea original en este aspecto. Consulte los manuales recomendados para las distintas materias.

Haga referencia a los *antecedentes* del problema. En los primeros cursos es suficiente con lo que encontrará en los textos.

Justifique *por qué* tiene sentido *hacer el trabajo.* Como lo normal en una práctica es replicar algún efecto bien conocido y su sentido puede parecer un poco superfluo, puede justificar, por ejemplo, que no existe suficiente evidencia.

Explique *cómo* va *a dar respuesta* al problema planteado.

La introducción conviene terminarla con una *deducción lógica* desde los planteamientos presentados hacia los resultados que *se esperan* obtener. Si lo desea, lo puede hacer en forma de *hipótesis.* Cuando vaya a estudiar más de una variable independiente, no olvide realizar una predicción por variable.

Método

El método ha sido el objeto de este texto, así que procure que se vea convenientemente reflejado lo mucho que ha aprendido. Con una correcta descripción del método otra persona debería *poder hacer otro trabajo* exactamente *igual* al suyo, de manera que un posible criterio de validez es dar a leer su método a otro compañero que no haya participado en su trabajo y pedirle que trate de repetirlo. Si se atasca en algún paso es que su redacción no era todo lo buena que parecía.

El método también sirve para que evalúen la calidad de su trabajo con los baremos científicos que ya conoce: *fiabilidad y validez.* Sea cuidadoso.

Existe una convención para organizar el método en varios subapartados. Siguiéndolos se asegura que no se le olvida nada de lo fundamental. Se los presentamos a continuación:

- *Sujetos*. En este apartado se debe especificar: quiénes, cuántos, de qué población, cómo se seleccionaron, cuál es el rango de su edad y la composición por sexos, si se les gratificó de alguna manera. En caso de una selección previa, puede contarse aquí o en el procedimiento.

- *Diseño.* Este subapartado puede ir en solitario o en conjunto con el procedimiento. Nosotros le recomendamos que lo haga siempre aparte, como forma de reflexión sobre el plan de investigación utilizado.

  Comience diciendo si ha sido un estudio descriptivo, un experimento o un cuasi experimento. A continuación si han utilizado grupos o un solo sujeto.

  Especifique cómo se ha/n operativizado la/s variable/s independiente/s y la dependiente.

  Después nombre el diseño específico utilizado. Diga cómo se ha administrado cada variable respecto a los sujetos (inter-grupos o intra-grupos), cuántos han sido los niveles de la/s VI(s) utilizados y cuáles los nombres de estos niveles.

- *Aparatos-materiales.* Cuando los instrumentos utilizados para llevar a cabo el trabajo tienen algún carácter especial o novedoso, descríbalos de forma independiente. Si no es el caso, si son materiales bien conocidos, integre los aparatos o materiales en la narración del procedimiento.

- *Procedimiento.* El procedimiento parece la parte más fácil del informe porque consiste simplemente en contar cómo se hicieron las cosas. Sin embargo, es muy difícil hacerlo bien. Por un lado está la tentación de ser demasiado meticuloso, si se tratase de las instrucciones de funcionamiento de algunos aparatos: "para poner en marcha el magnetófono desplace la palanca de la posición de *off* a la de *on",* o justamente la tentación contraria: darlo todo por sabido, lo que hará que el lector quede desconcertado; algo así como la frase: "tras presentar los estímulos, medimos la memoria de los sujetos".

  No espere hacerlo bien en la primera redacción. Una buena estrategia es la siguiente: haga un borrador, déselo a un amigo y haga que éste le explique a un tercero cómo se hizo el trabajo, estando usted delante. Pasarán un buen rato y podrá comprobar cómo lo que ha entendido su amigo difiere de lo que usted realmente hizo.

Resultados

En esta sección se presenta el *resumen* de los datos obtenidos y los cálculos estadísticos que se han hecho con ellos.

No olvide que se trata de un resumen de los datos, no tiene que grapar las hojas de respuesta de todos los sujetos, ni siquiera las matrices de sujetos por condiciones o ensayos.

Estos serían considerados los datos en bruto. Lo que usted tiene que incluir son los estadísticos que resumen los datos en bruto. Estos estadísticos son, por ejemplo, los índices de tendencia central y de dispersión.

Como no siempre se utilizan los mismos estadísticos, no dé por hecho que está claro de qué medidas se trata. Especifique, antes de informar de los valores, qué estadísticos son los que ha utilizado. Por ejemplo: "a continuación presentamos las medias de los dos grupos. El grupo de, "choque emocional" obtuvo un valor de 10 y el grupo, sin choque emocional, de 15".

Después de presentar los estadísticos descriptivos, presente los estadísticos de contraste. Aunque usted no lo pueda hacer todavía, sí que deberá ser capaz de identificarlos cuando lea un artículo . Por ejemplo: "la diferencia de medias fue estadísticamente significativa".

Tenga cuidado de no precipitarse al comentar los resultados. Es suficiente con decir en qué condiciones se obtuvieron valores más altos que en otras. Los comentarios en profundidad de los resultados se deben dejar para la discusión. Por ejemplo: "el grupo de choque emocional" recordó peor que el grupo "sin choque emocional".

Los resultados, a menudo, se presentan en tablas-resúmenes y en gráficas para mejorar la claridad expositiva.

Discusión

Comience con una frase que *resuma lo más relevante* del trabajo –no vuelva a repetir exactamente lo mismo que ya dijo en los resultados–. Siga *conectando* los resultados que ha obtenido con los que dijo en la introducción que esperaba encontrar. Si no coinciden, no se desespere. Eso significa que tiene que aguzar el ingenio y tratar de encontrar una explicación, normalmente metodológica, ya que es difícil que en una práctica se puedan mantener los controles de los experimentos originales llevados a cabo por expertos en laboratorios.

Recuerde: en caso de encontrar los datos que esperaba, las *hipótesis nunca se confirman,* simplemente no se rechazan, o si quiere huir de la antipática frase

con la doble negación, puede decir que los datos permiten mantener la hipótesis, o que son congruentes con la hipótesis planteada; en caso de no encontrar los datos esperados es más fácil, las hipótesis se rechazan, sin más.

Después de ligar los resultados con sus hipótesis, debe *conectar* este hecho con los datos de las *otras investigaciones* a la que se hacía mención en la introducción.

Es una buena costumbre *criticar* algunos de los fallos que usted mismo encuentra en su trabajo, una vez escrito. Una buena autocrítica indicará que usted ha aprendido haciéndolo, o sea, ha alcanzado el objetivo de la práctica.

En los artículos, es frecuente que los autores indiquen hacia donde se debería dirigir la próxima investigación. Usted puede indicar cómo mejorará su trabajo, si tuviera que repetirlo.

Referencias

Por referencias se conoce el listado de los autores y sus publicaciones, que se han hecho mención en el informe.

La Bibliografía, sin embargo, es un listado exhaustivo sobre las publicaciones de un tema, independientemente de que se hayan citado en el cuerpo del texto.

Los casos más comunes de referencias son: un libro, un capítulo de un libro y un artículo de una revista. Mejor que explicar la forma de hacerlo vamos a poner un ejemplo de cada tipo, usando referencias de las que aparecen en el texto.

1. *Libro:*
   Arnau, J. (1984). *Diseños experimentales en psicología y educación* (Vol. II). México D.F., Trillas.
2. *Capítulo de libro:*
   Diges, M., Rubio, M. E. y Rodríguez, M. C. (1992); Eyewitness memory and time of day. En F. Lösel, T. Bliesener (eds.), *Psychology and law* (páginas 317-320). Berlín: de Gruyter.
3. *Artículo:*
   Atkinson, J. W. y Litwin, G. H. (1960). Achievement motive and test anxiety conceived as motive to approach success and motive to avoid failure. *Journal of Abnormal and Social Psychology, 60,* 52-63.

   Trillas y de Gruyter son los nombres de las editoriales de los dos primeros ejemplos y México D.F. y Berlín las ciudades donde se ubican las editoriales.

El número que aparece a continuación del nombre de la revista, en el tercer ejemplo, es el volumen de la misma. Va también en cursiva.

Apéndices

En los apéndices se hace figurar información que sería muy prolijo situar en el texto. Se utilizan, por ejemplo, para indicar cuáles fueron las listas de estímulos que se presentaron, nuevos tests, programas de ordenador originales y cruciales para el trabajo, etc. En general los editores desaconsejan utilizarlos, salvo gran necesidad.

Documentación

María Eugenia Romano, una magnífica profesora nuestra, ya fallecida, comenzaba muchas veces sus clases con el siguiente lamento: "¡…que tengo yo que emplear la hora de clase en contarles estas cosas que están magníficamente expuestas en los libros...!".

En numerosas ocasiones, años más tarde, hemos comprendido su punto de vista. Esta es precisamente una de esas ocasiones, ya que lo mejor que puede hacer usted, si quiere saber cuáles son las fuentes documentales en Psicología y cómo se utilizan, es leer el libro de nuestros compañeros José Eugenio Ortega y José Miguel Fernández Dols (1980) "Fuentes documentales en Psicología". Es un texto claro, ordenado y perfectamente asequible.

Sin embargo, le sorprendería saber que a un alumno de postgrado –un poco despistado– le recomendaríamos el mismo texto. Lógicamente, hay una gran diferencia entre su situación y la de un alumno ya licenciado, y en consecuencia debe haber formas diferentes de acceder al tema de la documentación. En el caso del alumno superior se espera que –con un perfecto nivel de inglés a nivel de lectura– utilice las fuentes documentales, las estudie y sea capaz de asimilar la información para poder diseñar una investigación. En el caso de un alumno de primer ciclo se espera que reconozca el tipo de fuente documental cuando es nombrada en las clases o cuando es citada en un texto y que tenga un esquema mental de las formas que existen de profundizar teóricamente en un tema.

Para empezar vamos a resumir los tipos de fuentes documentales que seguro conoce como fruto de trabajar este texto y de estudiar otras materias. Las referencias a autores y años que ha ido viendo a lo largo de los distintos capítulos corresponden casi mayoritariamente a artículos en *revistas científicas* (vg: Atkinson y Litwin, 1960) generales o especializadas por temas– y a *libros* (vg: Azorín y Sánchez-Crespo, 1986). También había algunas referencias *a tesis doctorales* (vg: Gil, 1991), *memorias de licenciatura* (vg: Blanco, 1985) o *comunicaciones* presentadas *a congresos* (vg: Montero y León, 1991) incluso algún *documento interno sin publicar* (vg: Botella, 1992) o a un *manuscrito en preparación* (vg: Sáenz y León). Comentamos a continuación sólo los más importantes.

Dentro de las revistas hay un tipo particular que es de obligatoria consulta una vez que el tema de investigación está centrado, son las *revistas de revisión,* como el "Annual Review of Psychology", donde un grupo de investigadores revisan lo publicado en un tema –desde la última revisión, lo ordenan, lo comentan y lo critican desde su punto de vista. Leer una revisión sin formación previa en el tema es una labor ingrata y frustrante (si no nos cree, haga la prueba). El alumno de primer ciclo debería ser capaz de entender los artículos seleccionados por sus profesores antes de avanzar más.

En cuanto a los libros, se habrá dado cuenta que algunos son exclusivamente escolares –llamados *libros de texto*–, es decir, son obras pensadas para la formación curricular y en muchos casos obedecen al desarrollo de los programas oficiales, de las asignaturas. Otros libros, llamados de *lecturas,* son agrupaciones de escritos de distintos autores, pero aglutinados bajo el criterio de alguien, el compilador que ha perfilado un criterio para cubrir la información en un área determinada.

Como usted imagina, hay libros que tratan los temas más extensamente de lo exigido en los estudios de licenciatura. Estos libros *manuales* o *monografías* sólo son recomendables en niveles superiores de formación. Probablemente, el alumno de primer ciclo tiene bastante tarea con asimilar la información contenida en los libros de texto recomendados.

Antes de comentarle algunas formas de documentación más específicas queremos señalarle –para el caso de que no las conozca ya– la existencia de dos fuentes de gran utilidad, fácil acceso y recomendables para estudiantes noveles: las *enciclopedias especializadas y* los *diccionarios especializados.* Las enciclopedias son costosos y voluminosos trabajos dirigidos por profesores de reconocido prestigio y que constituyen una excelente manera de buscar información adicional o explorar alguna cuestión que queda fuera del programa. Consulte en su biblioteca sobre las disponibles. Los diccionarios son útiles para fijar el uso de términos, necesidad que es más apremiante en las ciencias más jóvenes. Quizá el orden sería asegurarnos primero en el diccionario de que la palabra que conocemos responde a lo que queremos conocer y después acudir a la enciclopedia. En general, conviene hacer una llamada de atención sobre las traducciones que no siempre son tan fieles como quisiéramos.

Presentamos ahora una de las herramientas de mayor utilidad entre los investigadores de Psicología: *Psychological Abstracts.* Se trata de una *revista de resúmenes,* elaborada sobre más de 1000 revistas científicas de ámbito internacional, utilizando los resúmenes de los informes de investigación. En el caso de "Psychological Abstracts", ésta se publica cada seis meses y contiene, no sólo los resúmenes de los artículos con los que usted está ya familiarizado, sino información

de libros, capítulos de libros y otros tipos de documentación escrita. La mayor ventaja de estas revistas de resúmenes es poder tomar la decisión de si merece la pena localizar el trabajo original, dada la información que proporciona el resumen. Otra ventaja no menos importante es la de disponer de un sistema de ordenación por áreas de conocimiento, áreas que permiten focalizar la búsqueda. Asimismo, dispone de dos índices de descriptores, uno por materias –algo así como las páginas amarillas– y otro por autores.

Los usuarios avanzados cuentan además con un magnífico complemento: se trata del *Thesaurus* de la APA. Un tesauro es una especie de diccionario de términos especializados de una materia. Su objetivo no es definir, sino fijar cuáles son los términos que se deben utilizar para facilitar la normalización del lenguaje científico. En el tesauro citado –lógicamente en inglés– además de la relación de términos en orden alfabético, hay una sección que los agrupa por su proximidad conceptual y otra que indica cuáles son los términos equivalentes para realizar búsquedas por ordenador.

Recordamos que nadie suele empezar a estudiar un tema buscando artículos en los "Psychological Abstracts". Si a alguien se le ocurriese, descubriría que el volumen de lo publicado sólo en los últimos diez años, sobre el más específico de los temas, es tal que no acabaría de leerlo todo en el tiempo que tiene previsto para realizar su trabajo, y esto sin suponer que se ha asimilado convenientemente. Para que nos entienda, ¿se le ocurriría tomar su primera clase de windsurfing en mar abierto, con olas de tres metros y vientos fuertes? Y, sin embargo, en estas condiciones es donde mejor se lo pasan los expertos.

La posibilidad de almacenar la información en soportes magnéticos está facilitando la búsqueda de información a través de ordenadores personales. Éstos, con el uso de programas específicos y conectados a redes, hacen que la búsqueda bibliográfica sea cada vez más extensa, más focalizada y más rápida. En cualquier caso, es muy probable que en su Facultad o Universidad exista un centro de documentación donde le puedan informar cuando lo necesite.

ANÁLISIS, INTERPRETACIÓN Y COMUNICACIÓN DE LOS RESULTADOS

Donald Ary*

Una vez reunidos los datos el investigador analiza e interpreta cuidadosamente los resultados y por último redacta un informe del estudio. Este capítulo contiene una breve exposición sobre cada una de esas etapas.

Análisis de los datos

Lo primero que hace el investigador para analizar la información es consultar el proyecto para verificar los planes originales referentes a la presentación de datos y la realización del análisis estadístico. Después de hacer esto, elabora una estrategia que le permita organizar los datos en bruto y efectuar los cálculos necesarios.

En la actualidad, los proyectos de investigación que se refieren a la pedagogía y a otras ciencias del comportamiento se caracterizan por la complejidad y requieren saber trabajar con computadora. Con objeto de lograr precisión y ahorrar tiempo y energía, algunos investigadores aprovechan la computación electrónica para analizar sus datos. Por ello se ofrece una breve explicación de las computadoras y de su aplicación de las mismas y de su aplicación en el procesamiento de datos.

Uso de las computadoras en el procesamiento de datos

El acontecimiento más importante en la historia de la investigación fue la invención de la computadora. Este fenómeno ha ensanchado el ámbito de la investigación y ha permitido hacer estudios que de otra manera ni siquiera se intentarían. Dado que las computadoras analizan grandes cantidades de datos con rapidez y eficiencia, se pueden diseñar estudios sin preocuparse por él número de variables ni por la complejidad de los análisis que se necesitan. Antes de la aparición de las computadoras, los investigadores se abstenían de realizar estudios que

* Tomado de: ARY, Donald et al: *Introducción a la investigación pedagógica.* México, D.F. Mc Graw-Hill Interamericana,1989, pp. 356-375.

incluyeran muchas variables y sujetos, porque requería tiempo y esfuerzo tabular y analizar los datos. Las pruebas estadísticas de gran complejidad y los análisis de multivariables no se efectuaban sino rara vez. La computadora es capaz de procesar grandes cantidades de información y hacer intrincados análisis estadísticos con increíble rapidez y eficiencia.

Preparación de los datos para la computadora

Para que el dispositivo de entrada acepte los datos, éstos deben elaborarse de modo que permitan la tabulación estadística, el análisis y el almacenamiento. El procesamiento de la información puede consistir en la conversión de los datos o en darles una forma más manejable. Esta elaboración pasa por varias etapas.

En la primera se codifican los datos en una forma que resulte apropiada para el análisis por computadora. La codificación consiste en cambiar la forma verbal de los datos de las variables y darles la forma numérica que necesita la computadora. Por ejemplo, el sexo de un sujeto no puede colocarse dentro de una computadora mediante el uso de palabras, sino que se aplica un código numérico a la variable, en el que 1 representa a un sujeto de sexo masculino y 2 a uno de sexo femenino. Los códigos se emplean para datos nominales como sexo, raza, nivel socioeconómico, grado escolar, escolaridad, estado civil, religión, preferencia política, lugar de residencia. Los códigos numéricos pueden aplicarse también a datos ordinales o de intervalo cuando se desee obtener categorías discretas para el análisis. Por ejemplo, en vez de usar las edades de los maestros que se incluyen en un estudio se las podría codificar con un 1 para los profesores menores de 25 años, con un 2 para aquéllos cuya edad fluctúa entre 26 y 30, con un 3 para los de 31 a 35, etc. La figura 1 contiene una serie de datos en que los tres primeros objetos fueron codificados y los restantes constituyen los datos efectivos. En un registro de datos los sujetos aparecen en los renglones y las variables, en columnas.

SERIE DE DATOS REGISTRADOS PARA UN ESTUDIO EXPERIMENTAL

Números de Códigos de los sujetos	Grupo	Sexo	Grupo escolar	Edad	Prueba de aptitudes académicas verbales	Prueba de aptitudes académicas matemáticas
1	1	1	3	16	450	495
1	1	2	4	17	560	460
2	2	1	3	15	570	530
2	2	2	4	18	620	640

Figura 1

Los expertos en las técnicas de programación pueden escribir sus propios programas. Por fortuna hay numerosos programas ya preparados, que se encuentran en la mayoría de los centros de computación. Se les denomina paquetes y son recomendables para quienes tienen muy poca o ninguna experiencia. Hay diferentes series de programas en paquetes. Una de las más conocidas es la Serie Biomédica, que fue creada para emplearse en los campos de la medicina y de la biología. Se dispone de gran número de tales programas para diversos análisis estadísticos. El Paquete Estadístico para las Ciencias Sociales es una de las series más populares en la investigación educacional. Contiene muchos de los procedimientos estadísticos de que frecuentemente se valen los sociólogos, su lenguaje es fácil de aprender además no requiere un conocimiento sobre el funcionamiento de las computadoras. Esos programas incluyen, entre otras cosas:

Mediciones codescriptivas de la tendencia central y de la dispersión.

Distribuciones indiferenciales de frecuencias.

Ji cuadrada (x^2).

Análisis de correlación bivariable y diagramas de dispersión.

Comparación de las medias de la muestra: pruebas t.

Análisis de varianza y de covarianza unidireccional con dirección n.

Correlación múltiple y regresión.

Correlación parcial.

Valoración escalar de Guttman.

Análisis factorial.

Por supuesto el uso de paquetes de programas no quita al investigador la responsabilidad de entender los principios estadísticos en que se apoyan, pues debe interpretar la salida del programa de computadora o hacer que un estadístico la explique.

Si no se cometen errores en la programación, los resultados pueden recopilarse en el registro de la computadora. Sin embargo, un solo error al disponer las tarjetas, la programación o al preparar las tarjetas de control impedirá que logre realizar la operación. En ese caso, habrá que reconocer y tratar de encontrar el problema que ocasionó la falla.

Conviene hacer una advertencia. La precisión de los resultados impresos depende totalmente de la información con que se alimente a la computadora. Esta no comete errores de cálculo, pero si recibe los errores de programación en que incurre el investigador, obedecerá las instrucciones en forma ciega pero eficiente y no producirá

más que "basura". Hay dos clases de errores que pueden cometerse en la programación. Los de lenguaje, entre ellos las instrucciones mal escritas, serán detectados por el aparato y entonces no aceptará la tarea. Pero como sí acepta los errores de lógica, tales como las instrucciones incorrectas, se obtendrá una producción muy cara de información sin significado.

Para terminar queremos hacer dos recomendaciones en relación con el empleo de las computadoras:

1. Aproveche al máximo las computadoras en el procesamiento de sus datos, pero realice una evaluación crítica del producto final. Estos aparatos no cometen errores pero tampoco corrigen los errores hechos en la programación o en la preparación de los datos para su procesamiento.
2. No utilice ciegamente cualquier programa que encuentre en los manuales. Usted debe decidir el tipo de procedimiento estadístico más apropiado para los datos que dispone.

PRINCIPIOS DE LA INTERPRETACIÓN

Una vez recabados los datos de investigación y terminado el análisis estadístico, el investigador comienza la fascinante tarea de interpretar los resultados. El principal objetivo de su trabajo es aportar algo nuevo al saber humano. Al llegar a la etapa de la interpretación, puede demostrar los conocimientos que se han adquirido en el proyecto y cómo encajan dentro del cuerpo general de la ciencia.

Cómo el proyecto facilita la interpretación

La base adecuada para interpretar los resultados de un estudio ha de ser puesta sistemáticamente a lo largo de las etapas de la realización del proyecto, incluso antes de comenzar el estudio. Al efectuarlo el investigador tiene presente en qué consistirán los datos y lo que les pueden decir, de ese modo se preparan para interpretarlos y hacerlos encajar en el acervo de conocimientos, de su disciplina.

Se supone que un plan cuidadosamente elaborado y presentado como un proyecto completo, producirá resultados cuya interpretación no será difícil y aportará información significativa si el estudio ha sido organizado de tal manera que las consecuencias de las hipótesis se expresen en observaciones confiables, la interpretación y el valor de las observaciones será obvio.

Importancia de apegarse al plan original

Cuando la propuesta haya sido aceptada y se ponga en marcha el proyecto, el estudio habrá de efectuarse sin apartarse en nada de lo planeado. Esta regla tiene consecuencias éticas y prácticas.

Para ilustrar el primer tipo de consecuencias supongamos que el señor Williams, un maestro de lengua extranjera, inventó con un gasto enorme de tiempo y esfuerzo un sistema para enseñar francés que él considera muy superior a los demás. Con el fin de probar la eficacia de su estrategia, establece un grupo experimental al que instruye por su método y un grupo de control que sigue otro. El profesor diseña una serie de pruebas semanales de aprovechamiento del francés que fungirá de variable dependiente. Imaginemos que durante las primeras semanas descubre que las puntuaciones medias del test son casi idénticas en los dos grupos. Como su método le costó mucho trabajo, le resulta difícil creer que no sea mejor que el otro y decide examinar a los dos grupos para indagar "qué salió mal". Descubre que el grupo experimental parece mostrar mayor conocimiento y apreciación de la vida y la cultura francesa. Deseoso como está por encontrar una diferencia entre las puntuaciones, opta por cambiar su variable dependiente por las puntuaciones obtenidas en las pruebas sobre la vida y la cultura francesas.

Este cambio será una trasgresión de las normas éticas. En dos grupos aleatorios siempre es posible encontrar una superioridad si se los somete a una atenta observación. Si el grupo experimental no hubiera parecido superior en el interés por la vida y cultura francesa, quizá habría sido superior en gramática inglesa o latina, en comportamiento o en cualquier otra variable. El maestro de idiomas tiene la obligación de efectuar el experimento tal como lo planeó en vez de modificar su variable dependiente después de observar una diferencia. Le es lícito anotar en algún lugar de su informe que el grupo experimental parecía superior en la comprensión de la vida y la cultura francesas, y que valdría la pena utilizar esa variable en otro experimento con otro grupo de sujetos. No es ético abandonar las variables independientes o dependientes que no parezcan "funcionar", ni agregar algunas nuevas y prometedoras. Tales cambios han de dejarse para futuros estudios.

La introducción de nuevas variables es también desaconsejable desde un punto de vista práctico, pues puede confundir los resultados de un estudio y obscurecer su significado. A menudo no siente la tentación de agregar nuevas variables interesantes que aparecen en el estudio. Pero la base teórica para interpretarlas todavía no ha sido puesta, por lo cual se aconseja dejarlas para otra ocasión.

Interpretación de los resultados esperados

Es comprensible que los investigadores se sientan complacidos cuando los resultados se ajustan a la estructura construida previamente y cuando la interpretación se efectúa según lo previsto. Se dice entonces que el estudio ha "funcionado" y que existe concordancia entre lo esperado y los resultados.

En tal caso sólo habrá que atender a unas pocas precauciones.

1. No hacer interpretaciones que excedan a la información. Esto puede parecer un precepto obvio, pero los investigadores se emocionan tanto al encontrar los resultados previstos que sacan conclusiones que no tienen una base válida en los datos. Incluso en las investigaciones publicadas es frecuente tropezarse con más interpretaciones de las que justifican los datos.

2. No olvidar las limitaciones del estudio. Por supuesto deberán de haberse identificado previamente en el estudio –las que se deben a una imperfecta confiabilidad y validez de los instrumentos, las ocasionadas por muestreo restringido, los problemas de validez de los instrumentos, es muestreo restringido, los problemas de validez interna, etcétera.

3. La ética requiere que el investigador comunique los problemas de validez interna que pudieran explicar los resultados. Si a pesar de que hizo todo lo posible por igualar las condiciones, las variables no experimentales fueron muy favorables para el grupo experimental y las del grupo de control fueron demasiado perjudiciales, debe darlas a conocer y tomarlas en cuenta al interpretar los resultados. (Por ejemplo, a pesar de la asignación aleatoria los maestros a los grupos, al grupo experimental tal vez le hayan tocado maestros expertos en su gran mayoría y al de control, principiantes.)

4. Recuérdese que la significación estadística sólo quiere decir que si hay grados de libertad idóneos es poco probable que los resultados dependan de la casualidad.

No indica en absoluto que los resultados sean significativos en la acepción general de la palabra –es decir, importantes y significativos o trascendentales. No piense que la significación estadística garantice que los resultados son de suma importancia.

Supongamos que dos grupos equivalentes fueron sometidos a dos diferentes sistemas de aprendizaje de ortografía durante dos años. Durante el experimento los sujetos que utilizan el sistema A muestran una ganancia media 2,15 años de crecimiento en pruebas estandarizadas, mientras que los que en el sistema B presentan una ganancia de 2,20. Si los grupos son grandes o si las diferencias dentro de los grupos son pequeñas, las diferencias entre medias tendrán significación estadística. Pero una diferencia de medio mes en un periodo de dos años es relativamente insignificante en términos prácticos. Si el sistema B requiere más tiempo por parte del alumno y del maestro o más materiales, a los profesores no les convendrían adoptarlo sólo porque según la estadística produce una ganancia mayor que el sistema A.

Pero si resultara menos costoso tendrían a preferirlo ya que en la práctica sus resultados son muy similares a los del otro sistema.

La posible importancia o intrascendencia de los resultados debe mencionarse en el proyecto antes de iniciar el estudio. Una investigación carece de valor si no proporciona información importante que enriquezca el saber grande que sea la significación estadística de los resultados.

Interpretación de resultados negativos

Los investigadores que descubren resultados contrarios a los que señalan las hipótesis suelen percatarse súbitamente de las deficiencias de su trabajo. La interpretación de los resultados se asemeja a una confesión. Los instrumentos fueron inadecuados para medir las variables; la muestra era demasiado pequeña y tenía tan poca representatividad que los resultados no pueden generalizarse válidamente a una población de objetivo extensa, etc. La percepción retrospectiva revela problemas de validez interna que explican por qué el estudio no corresponde a las expectativas.

Cualquiera de esas cosas puede haber ocurrido, pero las deficiencias deberán informarse sin importar los resultados. No obstante, la investigación es siempre una aventura que nos lleva a lo desconocido, de modo que no existe un "deberá ser" absoluto, el investigador predice los resultados de un estudio a partir de la teoría, la deducción y los resultados de análisis anteriores. Si éstos son tan concluyentes que no dejan ninguna duda acerca de los resultados, el estudio será útil.

Al emprender un estudio se afirma implícitamente que el resultado es una cuestión de conjetura, no de certidumbre. Al terminar el proyecto se entiende que los investigadores buscarán determinar imparcialmente la situación verdadera con los mejores instrumentos y procedimientos de que disponen. Por tanto, están obligados a aceptar e interpretar sus datos sin importar la relación, que guarda con la hipótesis. Cuando los resultados contradigan lo sostenido por la teoría en que se funda el estudio, la discusión del informe deberá incluir una reconsideración de la teoría original a la luz de los nuevos descubrimientos.

Los investigadores se muestran muy renuentes a presentar e interpretar los datos que se oponen a las investigaciones anteriores o a teorías ya aceptadas. Pero no debe excluir la posibilidad de que los resultados sean correctos y que los de estudios anteriores sean erróneos. El progreso de la pedagogía se retrasaría si los investigadores se negaran a comunicar los hallazgos que no concuerden con estudios previos. Los resultados contradictorios indican que una cuestión no está resuelta y pueden estimular otros análisis. La investigación ulterior o la elaboración de teorías puede conciliar a la postre a 1os resultados que parezcan contradictorios. La teoría es de carácter provisional y no debe disuadir para abstenerse de ofrecer una interpretación directa de lo que descubrió.

La reconsideración del fundamento teórico de un estudio pertenece a la parte de la discusión. No hay por qué retroceder y volver a escribir las secciones del informe dedicadas a la hipótesis y a la literatura sobre el tema del trabajo.

Interpretación de los resultados cuando se retiene la hipótesis de nulidad

Una hipótesis nula puede aceptarse por diversas razones, por lo cual la interpretación de ese hecho puede ser muy difícil. Una hipótesis de nulidad puede retenerse por tres motivos: 1) Realmente es verdadera. Tal vez no hay ninguna relación entre las variables, es posible que el tratamiento experimental no sea eficaz que el de control. 2) La hipótesis es falsa, pero los problemas de la validez interna contaminan la investigación a tal grado que no se pueda observar la relación entre las variables. 3) La hipótesis es falsa, pero la investigación carece del poder para rechazarla.

Cuando el investigador no sabe cuál de estas causas es la que interviene, no deberá valerse de ninguna de ellas para explicar los resultados.

Es incorrecto presentar la retención de una hipótesis de nulidad como prueba a favor de ausencia de relación entre las variables, pues se la debe interpretar como falta de pruebas de la veracidad o falsedad de la hipótesis. El comercial de una conocida pasta dental afirma que las pruebas demuestran que un dentífrico particular es insuperable para combatir las caries dentales. Si interpretamos el adjetivo insuperable en el sentido de ninguna diferencia significativa, podemos imaginar una prueba en que participó un pequeño número de sujetos o en que hubo numerosos problemas de validez interna.

Si lo que se desea en un experimento es retener la hipótesis de nulidad, la obtención de ese resultado es la cosa más sencilla.

Si se está estudiando una población pequeña y es posible realizar un censo completo de la misma, la retención de una hipótesis de nulidad puede interpretarse legítimamente como una carencia de relación entre las variables dentro de esa población. También adquiere credibilidad cuando interviene una muestra extensa. Por ejemplo, el informe de Coleman con más de 600,000 sujetos proporciona una base tan amplia que estamos dispuestos a admitir una falta observada de relaciones entre las variables como prueba de la inexistencia de relaciones. Sin embargo, en la mayoría de los estudios la retención de la hipótesis de nulidad no indica más que falta de pruebas.

Existe el peligro de que los investigadores se enamoran demasiado de sus hipótesis experimentales y cedan a la tentación de ignorar el hecho de que se retuvo la hipótesis de nulidad. Citan problemas de validez interna y declaran que los resultados habrían sido significativos si problemas imprevistos no hubiesen arruinado

el experimento. Desde luego necesitan comunicar todos los problemas de validez interna que surjan en un estudio, pero no deben usarlas como pretexto para justificar los resultados decepcionantes. Es preferible sugerir una investigación ulterior, planeada de manera que evite los problemas de validez interna, pero incluso en tal caso subsiste la obligación de mencionar que la retención de la hipótesis indica ausencia de pruebas y nada más.

El poder de un experimento designa la capacidad estadística de rechazar una hipótesis de nulidad cuando sea falsa. Este poder depende del tamaño de la muestra, la heterogeneidad de los sujetos con respecto a la variable dependiente, la confiabilidad de los instrumentos de medición y la naturaleza de los procedimientos estadísticos empleados para probar la hipótesis. Los investigadores deberán tomar en cuenta estos factores al planear un experimento. Algunos textos de estadística explican cómo planear los experimentos para que las relaciones significativas sean expresadas como observaciones estadísticamente significativas. Al planificar el estudio, hay que tomar en cuenta el poder del experimento; no se le ha de presentar al final para disculpar la carencia de significación estadística. Por ejemplo, no es lícito decir: "los resultados habrían tenido significación estadística si la muestra hubiese sido más grande".

Salvo raras excepciones, la única interpretación legítima de la retención de una hipótesis de nulidad es que constituye la prueba a favor de una conclusión que no ha sido observada.

Interpretación de relaciones no contenidas en las hipótesis

Se señaló antes que el investigador no deberá abandonar una hipótesis para seguir otras pistas más prometedoras que aparezcan durante su estudio. Esto no significa que han de ignorarse las relaciones que no figuren en la hipótesis y que puedan observarse al efectuar el estudio. Por el contrario, registran y analizan con el mismo rigor con que se analizan las contenidas en la hipótesis.

No obstante, tales hallazgos deberán tomarse con mayor desconfianza que los relacionados directamente con la hipótesis, ya que existe una posibilidad considerable de que en un estudio aparezcan relaciones falsas no contenidas en la hipótesis. Deben figurar en el informe, pero se las juzgará incidentales en relación con el fin principal de la investigación.

Habrán de ser tema de un estudio diseñado especialmente para examinarlas y sólo entonces servirán de base a las conclusiones.

EL INFORME DE INVESTIGACIÓN

Los resultados de un proyecto tienen poco valor si no se puede comunicar a otros. En consecuencia, es indispensable que los investigadores conozcan los procedimientos para redactar un informe. En esta sección se ofrece guía general para organizarlos y presentarlos.

El investigador comunica en su informe los procedimientos y los resultados del estudio, y expone las consecuencias de estas últimas y su relación con otros conocimientos sobre el tema.

Como los lectores son profesionales muy atareados, el informe debe ser muy conciso y contar con una estructura lógica. Las anécdotas, las descripciones de experiencias personales, las argumentaciones, y recursos ya no tienen cabida en él. Esto no quiere decir que el informe haya de ser aburrido y pedante. El estudio con que se haya abordado el tema se reflejará en la redacción.

El objetivo del informe consiste en presentar la investigación y no la personalidad del autor; por eso el tono ha de ser impersonal y nunca se emplea la primera persona. No se escribirá "Asigne aleatoriamente a los sujetos a dos grupos de tratamiento" sino "Los sujetos fueron asignados aleatoriamente a grupos de tratamiento".

A pesar del entusiasmo que produce la importancia de la labor, el autor no deberá jactarse de ello sino dejar esta evaluación en manos de los lectores y de la posteridad.

Existe un método formal y uniforme para presentar informes. Aunque a primera vista se piense que eso coacta la libertad, en la práctica cumple una función útil. Es importante disponer los informes de manera que los lectores sepan exactamente dónde encontrar lo que buscan. En el capítulo 3 se recomendó que, cuando se vayan a reunir los estudios que tratan del tema, lo primero que se hace es leer el resumen de un artículo de revista. Si el artículo carece de esa sección, no se puede evitar perder tiempo leyéndolo íntegro para saber si es de interés.

La existencia de un formato oficial elimina la necesidad de diseñar el propio. Al ahondar en este tópico se notará que el formato sigue las etapas un proyecto de investigación delineada en capítulos anteriores.

Un informe adopta las siguientes modalidades: 1) una tesis o disertación, 2) un artículo de revista, 3) una conferencia. Cada uno requiere un método propio.

Tesis

La mayoría de las universidades disponen de un manual que describe pormenorizadamente el formato que piden. Una vez elegido un manual, el informe se redacta de acuerdo con sus recomendaciones.

A continuación aparece el orden y los componentes generales descritos en la mayoría de los manuales de estilo:

I. Páginas preliminares

A) Portada

B) Página de aceptación

C) Prefacio o testimonios de gratitud

D) Índice

E) Lista de cuadros

F) Lista de figuras

II. Texto

A) Parte introductoria

1. Planteamiento del problema y justificación del estudio

2. Objetivos

3. Definición de términos

4. Literatura sobre el tema

B) Parte de los métodos y los resultados

1. Sujetos

2. Procedimientos

3. Instrumentos

4. Presentación y análisis de los datos

C) Exposición de los resultados

1. Interpretación

2. Consecuencias

3. Aplicaciones

D) Conclusiones y resumen

1. Conclusiones

2. Resumen

III. Páginas suplementarias

A) Bibliografía

B) Apéndices

C) Pequeño resumen autobiográfico (si se requiere)

D) Resumen

I. Páginas preliminares

Para preparar las páginas preliminares basta seguir fielmente las reglas de los manuales de estilo. No obstante, un aspecto de esas páginas que requiere una explicación especial es el título del estudio.

El título deberá describir, con la mayor brevedad posible, la naturaleza específica del estudio. Por ejemplo, consideremos: a) un estudio sobre niños con desventajas culturales en que se compara la celeridad de lectura de los alumnos que han participado en un programa del Head Start Project (proyecto de ventaja inicial) con los de un grupo apareado de niños sin experiencia preescolar formal, y b) el título "Una comparación de las puntuación obtenidas en una prueba de celeridad de lectura de niños con desventajas que han asistido a clases de ese programa durante seis semanas o más y de niños similares con experiencia preescolar". Aunque este título cubre el objetivo del estudio, resulta demasiado largo. Las frases como "comparación de... "un estudio de", "una investigación de" son a menudo superfluas. Además, la mayoría de los lectores sabrán que el proyecto Head Start es una experiencia cultural diseñada para niños con desventajas culturales. Sin embargo, irse al extremo opuesto y ofrecer un título demasiado corto o vago para que *cubra* la naturaleza del estudio constituye un error mucho más grave. Los títulos vagos o demasiado breves hacen necesario revisar el artículo para averiguar de qué se trata. Los títulos como "Ventaja inicial y celeridad" o "Lectura entre los niños con desventajas" ilustran esta deficiencia. El título deberá identificar las variables más importantes y las poblaciones de interés. Las definiciones operacionales de las variables más importantes y la descripción de las muestras no necesitan ir en el título.

El título correcto asegurará una posición adecuada en el índice, por lo cual una estrategia muy útil consiste en que los investigadores decidan en primer lugar bajo qué palabras claves desean situar sus estudios, y que a partir de esto lleguen a un título conciso. En el ejemplo las palabras claves serían celeridad de lectura y proyecto de ventaja inicial. Por tanto, un título apropiado podría ser: "Celeridad de lectura en niños que participan en el proyecto Head Start y de los que no participan en él".

Este título es razonablemente corto y sin embargo da una indicación válida al lector sobre la naturaleza del estudio.

Los títulos que llevan una carga emocional como "Tenemos que ampliar programas de la ventaja inicial" o "No permita que los niños con desvenas culturales se conviertan en lectores deficientes", deberán evitarse a toda costa. Tales títulos no ofrecen resultado de investigación, sino artículos teóricos e intentan vender un punto de vista.

II. Texto

A) Parte introductoria. Incluye todo lo que sucedió al colocar los cimientos de la investigación. Por lo general consta de materiales ya preparados para el proyecto, con pequeñas alteraciones. La formulación del problema y la justificación del estudio permanecen iguales, así como la enunciación de los objetivos, la definición de los términos y la revisión de los estudios sobre el tema. No obstante la relación breve de las fuentes de los datos y de los métodos se escribe en pretérito en lugar del futuro que se usa en el proyecto original. La revisión de los estudios relacionados con el tema suele presentarse como un capítulo independiente de esta parte.

B) Métodos y resultados. Esta parte del informe contiene cuatro categorías de información:

1. Sujetos. Deberá presentarse una clara descripción de la muestra, de preferencia al inicio del informe. Esto le permite al lector juzgar la potencial validez externa del estudio. Se define la población de la que se sacó la muestra y se especifica el método de muestreo. La clase de información que se ofrece en la descripción de la muestra cambiará de un estudio a otro, pero casi es posible determinar qué población incluir si se muestran las variables que podrían influir en las puntuaciones de criterio.

2. Procedimientos. El informe sobre los procedimientos del estudio deberá ser lo suficientemente completo para que cualquiera que desee reproducir la investigación encuentre toda la información. Una de las características del método científico es la posibilidad de confirmar los descubrimientos mediante la repetición de los procedimientos y la información observacional que requiere la reproducción. Esta sección comprende el diseño del estudio, el número de grupos (si el estudio es un experimento), el tratamiento de los sujetos y otra formación pertinente.

3. Instrumentos. Un informe de investigación deberá especificar todos los instrumentos de medición y los sistemas de observación adoptados en la realización del estudio. Estas descripciones pueden ser breves cuando se hayan usado mediciones corrientes como los exámenes de química del ACS-NSTA, puesto que en la bibliografía figurará la información pertinente sobre tales instrumentos. Si han sido creados instrumentos especiales para el estudio tiene que proporcionarse una descripción detallada de éstos, así como datos a favor de su confiabilidad y validez y una explicación de los procedimientos de puntuación.

4. Presentación y análisis de los datos. Una técnica conveniente para la presentación y el análisis estadístico de los datos consiste en organizar la

exposición a partir de las hipótesis, es decir, volver a enunciar la primera hipótesis en forma modificada y los resultados que se refieren a ella, y este procedimiento se repite en cada hipótesis.

Los cuadros y las figuras ayudan a presentar los datos con mayor claridad y precisión de la que sería posible si la misma información figurase en el texto. La mayoría de los manuales de estilo contienen ejemplos sobre los tipos comunes de cuadros y figuras e instrucciones para su construcción. Un cuadro bien integrado puede darle a1 lector una visión precisa de los datos.

Los cuadros elaborados durante un estudio casi nunca pueden incorporarse de modo directo en el informe. Por ejemplo, al finalizar un estudio, hay una lista alfabética de los sujetos y de sus puntuaciones en las mediciones de criterio. En lugar de insertar esta lista tal como está, convendría construirse un cuadro de la información en forma sumaria. Los cuadros de datos básicos en bruto pueden incluirse en el apéndice cuando se piense que contribuye a entenderles mejor. El primer cuadro del informe suele resumir los datos descriptivos como las medias, desviaciones estándar, correlaciones, porcentajes. En cuadros posteriores se muestran los resultados que se obtienen al aplicar a los datos la estadística inferencial y las pruebas de significación. Por ejemplo, podría usarse un cuadro sumario para presentar los resultados de un análisis de varianza.

Se aconseja disponer los cuadros en una forma que ilustre la relación de datos con las hipótesis del estudio. Los principiantes a menudo caen en tentación de incluir los datos en ambas formas, con lo cual hacen sus informes más largos y tediosos. Un método mejor consiste en exponer los datos mediante cuadros y figuras acompañadas de un texto que señale los descubrimientos más importantes e interesantes. Es muy importante relacionar la información de los cuadros con las hipótesis.

El fundamento estadístico del análisis de los resultados debe ser expresado con claridad. Es conveniente relacionar el tratamiento estadístico con la presentación de los datos.

C) Exposición de los resultados. Los resultados vuelven a interpretarse en relación con las hipótesis (o preguntas), y se comentan las aplicaciones y consecuencias del estudio.

1. Interpretación de los resultados. Esta parte del informe tal vez sea la más difícil pero también la más provechosa. La interpretación que investigador hace sobre los resultados los relaciona con la teoría, con los estudios sobre el tema y con los procedimientos de investigación.

2. Consecuencias. En esta parte deberá mencionarse también la contribución de los resultados al conocimiento en el campo general de estudio, el investigador explica cómo los resultados modificarán las teorías y sugiere futuros estudios que convendría efectuar.
3. Aplicaciones. Una afirmación sobre la aplicación de los resultados ayuda al lector a conocer el grado en que los descubrimientos pueden usarse en la práctica.

Las secciones dedicadas a las aplicaciones y consecuencias de los resultados a menudo no se exponen con suficiente amplitud porque se supone que también serán tan obvias para el lector. En realidad, durante el estudio el investigador logra una visión más profunda del problema que la que puedan tener los lectores. Por lo tanto, cabe esperar que sus interpretaciones sean más significativas que las de los lectores.

D) Conclusiones y resumen. Las secciones de las conclusiones y del resumen forman el corolario del informe.

1. Conclusión. La exposición de las conclusiones deberá limitarse a las que tienen un apoyo directo en los datos aportados por la investigación. Los investigadores tienden a sacar demasiadas conclusiones. Las hipótesis proporcionan una estructura conveniente para hacer conclusiones; es decir, los investigadores deberán indicar si los descubrimientos apoyan o no a las hipótesis. Es importante distinguir entre resultados y conclusiones. Un resultado es una observación directa. Una conclusión es una inferencia basada en los resultados. Por ejemplo, un estudio podría producir la siguiente observación: las puntuaciones medias en una prueba de ortografía de los estudiantes instruidos con el método A es significativamente más alta que la media de los que fueron instruidos con el método B. La conclusión de que el método A es más eficaz que el B no constituye un resultado directo del estudio sino que es una inferencia basada en los resultados.

   Los investigadores pueden incluir una breve exposición de sus ideas sobre las consecuencias de sus resultados y dar algunas recomendaciones para las posibles aplicaciones. También pueden sugerir nuevas preguntas de investigación que emergen de su estudio.
2. Resumen. Como será más leído que otras secciones del informe, su redacción debe ser muy clara y concisa. El resumen suele incluir otra reformulación sintética del problema o problemas, las características principales de los métodos y los hallazgos de mayor importancia. Después de hacer un borrador de esta sección, el autor deberá verificarla cuidadosamente para asegurarse

que el estudio y los resultados estén descritos en forma clara y concisa. Asimismo, será necesario verificar que no se haya introducido ninguna información que no figure en las partes anteriores. Es conveniente que un colega lea esta sección para tener la seguridad de que se está comunicando lo que se desea.

III. Páginas suplementarias

A) Bibliografía. Deben comprender todas las fuentes citadas en el texto y las notas a pie de página. Algunas universidades insisten en que sólo se enumeren éstas, pero otras solicitan que se citen además las obras pertinentes que no hayan sido mencionadas de manera específica. El manual de estilo seleccionado dará los detalles sobre el método de citar las obras. Es necesario seguir esas reglas al pie de la letra. De hecho, es recomendable aprendérselas de memoria antes de iniciar la búsqueda de la literatura para el proyecto. Si cada obra se cita en la forma correcta tal como aparece, se evita perder tiempo para encontrarlas de nuevo y disponerlas en la bibliografía. Es conveniente transcribir en tarjetas las indicaciones bibliográficas de manera que puedan colocarse en orden alfabético.

B) Apéndices. Los apéndices contienen materiales que no son lo suficientemente importantes para ponerlos en el cuerpo del informe, pero que puede tener valor para algunos lectores. Esos materiales abarcan entre otras cosas copias completas de las pruebas o cuestionarios ideados en la localidad, junto con las instrucciones y claves de puntuaciones de tales instrumentos, el análisis de los datos de las medidas utilizadas, las instrucciones verbales para los sujetos y los cuadros que sean muy extensos o que sean de importancia menor para el estudio.

C) Pequeño resumen autobiográfico de los autores de informes de investigación; datos académicos, experiencia, afiliación profesional y contribuciones anteriores.

D) Resumen. Muchas instituciones solicitan un resumen aparte de la disertación, el cual debe contener una formulación definitiva del problema y las descripciones concisas de los métodos de investigación, los resultados de mayor importancia, las conclusiones y las consecuencias. El resumen debe tener extensión limitada (por lo general 600 palabras o menos). Las páginas se numeran independientemente y se colocan al comienzo o al final de disertación.

Artículo de revista

Al preparar un artículo de investigación para publicarlo en una revista, un buen primer paso consiste en averiguar qué revistas publican la mayor cantidad de material acerca del tópico que se va a explorar. La información sobre el procedimiento para presentar artículos suele encontrarse en la parte interior de la portada de la revista.

Algunas especifican el manual de estilo que habrá de usarse, por ejemplo el Manual de Publicaciones de la Asociación Estadounidense de Psicología o el Manual de Estilo de la Asociación Nacional de Educación. Si no se señala ningún manual, el estilo preferido, método de citar obras y otros detalles se escogen consultando artículos del ejemplar reciente de la revista.

Un artículo de investigación sigue las mismas líneas generales que una disertación, pero debe ser más corto. La función de las tesis o disertaciones es demostrar la competencia de un estudiante, y se requiere una exposición completa de la investigación afín, descripción exhaustiva de los procedimientos, tabulación íntegra de los resultados y una elaboración reflexiva. El artículo de revista, en cambio, sólo exige que el autor comunique su contribución al conocimiento. Como la revista dispone de poco espacio y los lectores poco tiempo, el artículo debe ser conciso.

La sección de la literatura sobre el tema contiene sólo aquellos resultados y argumentos que forman la base del problema. La formulación general del problema se da en un párrafo o, de ser posible, se omite. En este último caso el artículo comenzará con la hipótesis. La parte destinada a procedimientos se presenta también con gran brevedad, si es posible se incluye toda la información necesaria para reproducir el experimento. La sección de los resultados es la de mayor interés para el lector, en consecuencia ocupará una proporción mayor del artículo que en una disertación. Sólo los hallazgos más importantes deberán exponerse en detalle.

Una breve carta explicativa deberá acompañar al manuscrito, junto con sellos de correo para su devolución si no es aceptado. El editor suele remitirle al autor una tarjeta en que acusa recibo del artículo y hace circular copias del mismo al comité de lectura para que lo revise. Transcurre un considerable lapso de tiempo antes que se informe al autor si su trabajo fue aceptado o rechazado (casi siempre 6 semanas). Si el artículo es aceptado, pasan muchos meses antes que aparezca en la revista.

Cuando una revista rechaza un manuscrito, la notificación suele ir acompañada de una explicación de las razones. Tal rechazo no quiere decir que no merezca publicarse. La decisión puede depender de muchos factores, a saber: la competencia por el espacio, los cambios de la política editorial o los prejuicios de los miembros del comité de lectura. Un artículo rechazado por una revista puede ser revisado y sometido a otra. Muchos artículos recorren diversas revistas hasta encontrar que finalmente son aceptados. No es correcto someter un artículo a varias revistas simultáneamente.

Conferencia

Algunos investigadores descubren que los ensayos que se leen en convenciones profesionales constituyen una buena manera de mantenerse al día en el campo. Ello se debe a que transcurre un gran lapso entre la terminación de un proyecto de investigación y su publicación. El periodo es a veces tan largo las revistas especializadas han sido motejadas de archiveros.

Las ponencias que se presentan en las reuniones profesionales pasan por preparación muy similar a la de los artículos de revista. No siempre irán de ser informes de investigaciones terminadas, sino que pueden constituir informes sobre los progresos de proyectos en curso. El ensayo es menos formal que un artículo de revista y por lo general puede ajustarse con mayor precisión a su audiencia. Por lo común se piensa que el auditorio esté familiarizado con los detalles de las investigaciones afines y con los métodos de medición.

La ponencia se organiza en la siguiente forma:

1. Formulación directa de la hipótesis.
2. Breve descripción de los procedimientos.
3. Resultados, conclusiones y consecuencias.

El tiempo permitido por leer una ponencia suele ser bastante corto, usualmente menos de quince minutos. En consecuencia, deberá abordar los aspectos más importantes del estudio. Un método práctico consiste en asignar de 2 1/2 a 3 minutos por página del tamaño de carta mecanografiada a doble espacio.

Cuando los cuadros y figuras ayuden a la presentación, conviene distribuirse copias entre la audiencia. Algunos oradores también distribuyen copias del texto de su ponencia entre los oyentes pero como la audiencia la lee en silencio con mayor rapidez que el conferenciante, la atención puede perderse. Es preferible disponer de una descripción completa del estudio mimeografiada, que se entrega a quienes la soliciten después de la presentación de la ponencia.

ERRORES COMUNES EN LOS INFORMES DE INVESTIGACIÓN

Dvorak estudió cerca de 18,000 informes de educación comercial y compiló una larga lista de deficiencias que nos proporciona una buena lista de verificación. Los redactores de informes de investigación harán bien en estudiarla.

Título del informe de investigación. Demasiado largo. No refleja con exactitud el problema resuelto; es decir, es más amplio o estrecho que el problema planteado.

El problema. Se omiten la formulación y el análisis del problema, las delimitaciones, definiciones y objetivo(s) del estudio.

Todo ello se trata, tan a la ligera o está tan esparcido en el informe que resulta dificilísimo determinar el problema que el investigador intenta resolver.

La formulación y el análisis del problema, las delimitaciones, las definiciones y el objetivo(s) del estudio figuran en el informe de manera organizada pero se observan algunas deficiencias graves en los siguientes aspectos:

El problema es demasiado amplio como para que lo solucione una persona con recursos limitados.

La importancia del problema es nula o insignificante (se tiene la impresión de que la persona estaba interesada únicamente en satisfacer un requisito del curso y no en obtener una experiencia profesional de valor).

La formulación del problema es ambigua, adolece de un exceso de palabras, demasiado largo y complicado (contiene párrafos que podrían incluirse en las delimitaciones) y no está en armonía con los descubrimientos que comunica.

El análisis no logra redondear el concepto del problema; no se identifica o no se indica con claridad el acervo de datos necesarios como base para la solución del problema. La explicación de un problema bastante difícil de captar está mal escrita o fue omitida.

Se pasan por alto las delimitaciones más importantes que se requieren para fijar claramente el ámbito del estudio, no están colocadas donde beneficiaría al lector o no se expresan adecuadamente. Las delimitaciones afines no están situadas en proximidad una de otra.

No hay definiciones de los términos que se necesitan para una clara comprensión del estudio.

Literatura sobre el tema. Falta esta literatura.

Aunque hay informe de la literatura afín, pueden observarse ciertas deficiencias.

Algunos de los artículos presuntamente relacionados con el tema, en realidad no lo están en absoluto o sólo guardan un nexo remoto. Se tiene la impresión de que esta parte del informe fue "rellenada".

No hay informe sobre las relaciones de los artículos con el estudio.

Casi todo se reduce a una lista de sumarios; esto es, no se capta ninguna norma sólida de organización. En otras palabras, aunque la clasificación es posible no se intenta realizarla.

El volumen y la pertinencia de la literatura no se indican desde el comienzo; es necesario leer toda la sección y enjuiciarla uno mismo.

No se utiliza el original de un artículo a pesar de que sería relativamente fácil de obtener.

Ninguno de los artículos relacionados con el tema se acompaña de su respectiva información bibliográfica.

Métodos de investigación y el procedimiento. No se informa sobre los métodos de investigación ni sobre el procedimiento seguido.

Los métodos de investigación y el procedimiento están descritos en forma tan ambigua y esquematizada que la importancia de la investigación es difícil de captar.

Ambos aparecen en el informe, pero pueden observarse serias deficiencias:

La adecuación de los métodos de investigación es discutible.

Los métodos de investigación fueron identificados de modo incorrecto.

El procedimiento que se siguió está muy disperso; es decir, una falta de organización impide entender la totalidad del plan.

La naturaleza de los datos empleados en la solución del problema no se describe, o la descripción que constituye no es otra cosa que una repetición de los elementos del problema.

Las fuentes de actos no se nombran o resultan inadecuadas en términos del problema por resolver. No se da la fecha de los materiales publicados.

No existen indicaciones sobre las bases para preparar los instrumentos con que se recabaron los datos: en el informe no figuran copias de ellos. Los instrumentos no se confeccionaron con suficiente cuidado, ninguno de ellos ha pasado por una prueba de ensayo. Resultan inapropiados para recopilar el tipo de datos que se busca.

Las bases para la selección de los casos no se mencionan. El número de casos es muy pequeño o demasiado grande.

No se hacen pruebas para precisar el monto, la validez y la confiabilidad de los datos o se aplican otras pruebas injustificables.

Hay detalles innecesarios –esto es, no se discierne entre lo superfluo y lo importante– que hacen tediosa y difícil la lectura del procedimiento y obstaculizan la detección de los pasos mas importantes que se dan.

Se omiten las etapas para procesar e interpretar los datos y para realizar generalizaciones, o han sido elaboradas en forma tan deficiente que no se puede determinar fácilmente lo que hizo el investigador.

Los instrumentos estadísticos se usan de manera incorrecta.

En los estudios experimentales 1os factores, a diferencia de otras, no son controlados con cuidado.

Resultados. El informe de los resultados se basa en datos que no se explican en la formulación ni en el análisis de problemas ni tampoco en el procedimiento.

Aparecen datos sin procesar y la forma que se emplea es demasiado áspera para que el lector obtenga el máximo beneficio de los resultados. Puede decirse que los "datos" no se distinguen de los "hallazgos".

Los resultados están incompletos; es decir, no se explican todos los datos de los casos incluidos en el estudio.

Los prejuicios del investigador son patentes.

Los resultados secundarios reciben demasiada atención; esto es, no están enfocados desde la perspectiva adecuada.

Los resultados de los subgrupos no se comunican.

El formato para la presentación de los datos bloquea la penetración.

La interpretación de los datos se confunde con el resumen de los mismos.

El resumen es defectuoso en puntos estratégicos.

Generalizaciones. No se traza una distinción entre hallazgos y generalizaciones. Por ejemplo, a los resúmenes de los resultados se les califica de "conclusiones".

No se hacen generalizaciones, aunque las bases para efectuarlas estén claras. Se ejecutan generalizaciones más allá de lo que permiten los datos recopilados.

Las premisas para sacar la generalización son discutibles.

Los prejuicios del investigador son patentes.

Las generalizaciones están basadas en la literatura afín y no en los resultados del estudio.

No se distingue entre conclusiones y recomendaciones.

RESUMEN

Como casi todas las universidades poseen servicios de computación electrónica, los investigadores pueden utilizarlas en el procesamiento de sus datos. El uso de las computadoras ahorra tiempo y energía y garantiza un alto grado de precisión en los cálculos muy complejos.

Para emplear computadoras en el procesamiento de los datos, el investigador necesita traducirlos a una forma que pueda ser leída por ellas e impartirles

instrucciones por medio de un programa adecuado para que actúen sobre la información.

Se requiere un entrenamiento para escribir programas de computación, pero el investigador puede usar, en el análisis de sus datos, los programas ya preparados con que se cuenta actualmente.

La interpretación de los resultados de un estudio será una labor directa si, en el proyecto, el investigador ha puesto un fundamento apropiado pare el estudio. Ténganse presentes estas recomendaciones: 1) la interpretación habrá de basarse estrictamente en los datos derivados del estudio, 2) los problemas de validez interna y externa y otras limitaciones del estudio se tomaran en cuenta, y 3) Las conclusiones deben presentarse como afirmaciones de probabilidad y no como hechos.

Los resultados negativos merecen el mismo respeto e interpretación que los positivos. La retención de una hipótesis de nulidad no significa que los datos son insuficientes, y nada más. Los resultados que no se obtengan a partir de la hipótesis deben presentarse como fuentes de futuras hipótesis.

BIBLIOGRAFÍA

ANGERA, María Teresa, (1997): *Metodología de la observación en las ciencias humanas.* Madrid, Editorial Cátedra S.A.

ARNAU, Jaime, (1978): *Métodos de investigación en las ciencias humanas.* Barcelona, Editorial Omega.

ARY, JACOBS y RAZAVIEH, (1982): *Introducción a la investigación pedagógica.* México, Editorial Interamericana.

BOOTH, Wayne C. y otros, (2004): *Cómo convertirse en un hábil investigador.* Barcelona. Editorial Gedisa.

BRIONES, Guillermo, (1986): *Métodos y técnicas de investigación para las ciencias sociales.* México, Editorial Trillas.

BRIONES, Guillermo, (1986): *Curso avanzado de técnicas de investigación social aplicadas a la educación: Módulo I.*. Santiago de Chile, Editorial IDCR.

CALELLO, Hugo y NEUHAUS, Susana, (1999): *Método y antimétodo.* Buenos Aires, Ediciones Colihue.

CASTRO, Luis, (1977): *Diseños experimentales sin estadística.* México, Editorial Trillas.

CATALDI AMATRIAN, Roberto M., (2001): *Los informes científicos: Cómo elaborar tesis, monografías, artículos para publicar.* Buenos Aires, Editorial Lugar.

CATENA, Andrés y otros, (2003): *Análisis multivariado: Un manual para investigadores.* Madrid, Editorial Biblioteca Nueva.

COCHRAN, William y COX, Gertrude, (1985): *Diseños experimentales.* México, Editorial Trillas.

DAY, Robert, (1996): *Cómo escribir y publicar trabajos científicos.* Washington. OPS, Publicación Científica Nº 558.

De KELETE, Jean-Marie, (1995): *Metodología para la recogida de información.* Madrid, Editorial La Muralla.

ECHEVERRÍA, Javier, (2003): *Introducción a la metodología de la ciencia: La filosofía de la ciencia en el siglo XX.* Madrid, Editorial Cátedra.

ELLIOT, Jhon, (2000): *La investigación en Educación.* Madrid, Editorial Morata.

ESCOTET, Miguel A., (1980): *Diseño multivariado en psicología y educación.* Barcelona, Editorial CEAC.

FESTINGER, León y KATZ, Daniel, (1992): *Los métodos de investigación en las ciencias sociales.* Madrid, Editorial Paidós.

FEYERABEND, Paul K., (1981): *Contra el método.* Barcelona, Editorial Ariel.

FOX, David, (1981): *El proceso de la investigación en educación.* Pamplona, Ediciones de la Universidad de Navarra.

GLASS y STANLEY, (1985): *Métodos estadísticos aplicados a las Ciencias Sociales.* México, Editorial Prentice Hall International.

GOODE, William y HATT, Paul, (1975): *Métodos de investigación social.* México, Editorial Trillas.

GUTIÉRREZ PANTOJA, Gabriel, (1980): *Metodología de las ciencias sociales.* México, Editorial Harla.

HABER, André y RUNYON, Richard, (1973): *Estadística general.* México, Fondo Educativo Interamericano.

HERNÁNDEZ, FERNÁNDEZ y BAPTISTA, (1998): *Metodología de la Investigación.* México, Mc Graw-Hill.

KERLINGER, Fred, (1975): *Investigación del comportamiento.* México, Editorial Interamericana.

KERLINGER, Fred, (1981): *Enfoque conceptual de la investigación del comportamiento.* México, Editorial Interamericana.

KLIMOVSKY, Gregorio, (1997): *Las desventuras del conocimiento científico.* Buenos Aires, AZ Editora.

KHUN, Thomas S., (1997): *¿Qué son las revoluciones científicas y otros ensayos?* Barcelona, Editorial Paidós.

LEÓN y MONTERO, (1999): *Diseño de investigaciones.* Madrid, Mc Graw-Hill.

LIGHT, Richard y PILLEMER, David, (1984): *Revisando investigaciones.* Washington, OPS.

LUNA CASTILLO, Antonio, (1998): *Metodología de la tesis.* México, Editorial Trillas.

MARTÍN, Michael y Mc INTIRE, Lee, editores, (1994): *Readings in the philosophy of social science.* Cambridge, Publicaciones del Instituto Tecnológico de Massachussets.

Mc. KERNAN, James, (2001): *Investigación–acción y currículo.* Madrid, Editorial Morata.

MEJÍA, Elías y REYES, Edith, (1994): *Operacionalización de Variables Conductuales.* Lima, CENIT Editores.

MEJÍA, Elías, (1996): Factores del éxito académico en estudios de post grado. En: *Revista Peruana de Educación.* Año I Número 1. Enero 1996. Lima, Optimice, División Editorial.

MEJÍA, Elías, compilador, (2001): *La investigación científica.* Lima, Cenit Editores.

MELTZOFF, Julián, (2000): *Crítica a la investigación. Psicología y campos afines.* Madrid, Alianza Editorial.

MÉNDEZ, Carlos, (1998): *Metodología, Guía para elaborar diseños de investigación.* Bogotá, Mc Graw-Hill.

MÉNDEZ, GUERRERO, MORENO y SOSA, (1998): *El protocolo de investigación. Lineamientos para su elaboración y análisis.* México, Editorial Trillas.

MÉNDES RAMÍREZ, Ignacio y otros, (2000): *El protocolo de investigación: Lineamientos para su elaboración y análisis.* México, Editorial Trillas.

MORIN, Edgar, (1999): *El método.* Madrid, Ediciones Cátedra S. A.

MUÑOZ RAZO, Carlos, (1998): *Cómo elaborar y asesorar una investigación de tesis.* México, Editorial Prentice Hall.

NAGEL, Ernest, (1961): *La estructura de la ciencia.* Madrid, Editorial Paidós.

NAMKFOROOSH, (2003): *Metodología de la investigación.* México, Editorial Limusa.

PIERGIORGIO, Cobertta, (2003): *Metodología y técnicas de la investigación social.* Madrid, Editorial Mc Graw-Hill.

POLIT y HUNGLER, (1998): *Investigación científica en ciencias de la salud.* México, Mc Graw-Hill Interamericana.

POPPER, Karl, (1980): *La lógica de la investigación científica.* Madrid, Editorial Tecnos.

RAMOS, Manuel Miguel y otros, (2004): *Manual de métodos y técnicas de investigación en ciencias del comportamiento*. Madrid, Editorial Biblioteca Nueva.

RODRÍGUEZ, Aroldo, (1980): *Investigación experimental en psicología y educación.* México, Editorial Trillas.

RODRÍGUEZ GÓMEZ, Gregorio y otros; (1996): *Metodología de la investigación cualitativa.* Málaga, Editorial Aljibe.

ROJAS SORIANO, Raúl, (1986): *El proceso de la investigación científica*. México, Editorial Trillas.

ROSAS, Lucía, (1990): *Iniciación del método científico experimental.* México, Editorial Trillas.

RUIZ LABUENAGA, José y otros, (2002): *Cómo elaborar un proyecto de investigación social.* Bilbao, Editorial de la Universidad de Deusto.

SALKIND, Neil, (1998): *Métodos de investigación.* México, Prentice Hall.

SAMAJA, Juan, (1993): *Epistemología y metodología: Elementos para una teoría de la investigación científica.* Buenos Aires, Editorial Eudeba.

SCOTT, William A. y WERTHEIMER, M., (1981): *Introducción a la investigación de la Psicología.* México, Editorial El Manual Moderno.

SIEGEL y CASTELLAN, (1998): *Estadística no paramétrica.* México, Editorial Trillas.

SIERRA BRAVO, R., (1986): *Tesis doctorales y trabajos de investigación científica.* Madrid, Editorial Paraninfo.

SIERRA BRAVO, R., (1992): *Técnicas de investigación social, teoría y ejercicios.* Madrid, Editorial Paraninfo.

TABORGA, Huáscar, (1992): *Cómo hacer una tesis.* México, Editorial Mc Graw-Hill.

TAMAYO y TAMAYO, Mario, (2004): *El proceso de la investigación científica*. México, Editorial Limusa.

VALOR YEBENES, Juan A., (2000): *Metodología de la investigación.* Madrid, Editorial Biblioteca nueva. Madrid.

ZORRILLA, Santiago y otros, (1997): *Metodología de la investigación.* México, Editorial Mc Graw-Hill.

ZORRILLA y TORRES, Xamar, (1992): *Guía para elaborar la tesis.* México, Editorial Mc Graw-Hill.

CEPREDIM

Se terminó de imprimir
en el mes de noviembre de 2005,
en los talleres gráficos del
Centro de Producción Editorial e Imprenta de
la Universidad Nacional Mayor de San Marcos
Jr. Paruro 119. Lima 1.
Teléfono: 619-7000 anexos: 6009, 6011, 6015
Correo electrónico: cepedit@unmsm.edu.pe
Tiraje: 1000 ejemplares